

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
28 November 2002 (28.11.2002)

PCT

(10) International Publication Number
WO 02/094780 A2

(51) International Patent Classification⁷: **C07D**

(21) International Application Number: PCT/US02/16634

(22) International Filing Date: 23 May 2002 (23.05.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/293,665	24 May 2001 (24.05.2001)	US
60/298,712	15 June 2001 (15.06.2001)	US
60/303,418	6 July 2001 (06.07.2001)	US
60/306,967	19 July 2001 (19.07.2001)	US
60/308,183	27 July 2001 (27.07.2001)	US
60/343,007	19 December 2001 (19.12.2001)	US
60/357,675	15 February 2002 (15.02.2002)	US
60/376,988	30 April 2002 (30.04.2002)	US

(71) Applicant (*for all designated States except US*): **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **LU, Dyung Aina, M.** [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). **AZIMZAI, Yalda** [US/US]; 5518 Boulder Canyon Drive, Castro Valley, CA 94552 (US). **DING, Li** [CN/US]; 3353 Alma Street # 146, Palo Alto, CA 94306 (US). **LEE, Ernestine, A.** [US/US]; 20523 Crow Creek Road, Castro Valley, CA 94552 (US). **HAFALIA, April, J., A.** [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). **BECHA, Shanya, D.** [US/US]; 21062 Gary Drive # 117, Castro Valley, CA 94546 (US). **TANG, Y., Tom** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **LAL, Preeti, G.** [US/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). **GRIFFIN, Jennifer, A.** [US/US]; 33691 Mello Way, Fremont, CA 94555 (US). **GURURAJAN, Rajagopal** [IN/US]; 5591 Dent Avenue, San Jose, CA 95118 (US). **RAMKUMAR, Jayalaxmi** [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). **ELLIOTT, Vicki, S.** [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). **ARVIZU, Chandra, S.** [US/US]; 1706 Morocco Drive, San Jose, CA 95125 (US). **LUO, Wen** [CN/US]; 181 Cymbidium Circle, South San Francisco, CA 94080 (US). **SWARNAKAR, Anita**

[CA/US]; 8 Locksley Avenue #5D, San Francisco, CA 94122 (US). **DUGGAN, Brendan, M.** [AU/US]; 243 Buena Vista Avenue # 306, Sunnyvale, CA 94086 (US). **TRAN, Uyen, K.** [US/US]; 2638 Mabury Square, San Jose, CA 95133 (US). **WALIA, Narinder, K.** [US/US]; 890 Davis Street A # 205, San Leandro, CA 94577 (US). **GANDHI, Ameena, R.** [US/US]; 705 5th Avenue, San Francisco, CA 94118 (US). **YAO, Monique, G.** [US/US]; 1189 Woodgate Drive, Carmel, IN 46033 (US). **KHAN, Farrah, A.** [IN/US]; 9445 Harrison Street, Des Plaines, IL 60016 (US). **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). **BOROWSKY, Mark, L.** [US/US]; 122 Orchard Avenue, Redwood City, CA 94061 (US). **ZEBARJADIAN, Yeganeh** [IR/US]; 830 Junipero Serra Boulevard, San Francisco, CA 94127 (US). **RICHARDSON, Thomas, W.** [US/US]; 616 Canyon Road #107, Redwood City, CA 94062 (US). **MARQUIS, Joseph, P.** [US/US]; 4428 Lazy Lane, San Jose, CA 95135 (US). **CHIEN, David** [US/US]; 444 Yerba Buena Avenue, Los Altos, CA 94022 (US). **JIN, Pei** [US/US]; 320 Curtner Avenue #D, Palo Alto, CA 94306 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PI, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: KINASES AND PHOSPHATASES

(57) Abstract: The invention provides human kinases and phosphatases (KPP) and polynucleotides which identify and encode KPP. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of KPP.

WO 02/094780 A2

KINASES AND PHOSPHATASES

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of kinases and phosphatases
5 and to the use of these sequences in the diagnosis, treatment, and prevention of cardiovascular
diseases, immune system disorders, neurological disorders, disorders affecting growth and
development, lipid disorders, cell proliferative disorders, and cancers, and in the assessment of the
effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of
kinases and phosphatases.

10

BACKGROUND OF THE INVENTION

Reversible protein phosphorylation is the ubiquitous strategy used to control many of the
intracellular events in eukaryotic cells. It is estimated that more than ten percent of proteins active in
a typical mammalian cell are phosphorylated. Kinases catalyze the transfer of high-energy phosphate
15 groups from adenosine triphosphate (ATP) to target proteins on the hydroxyamino acid residues
serine, threonine, or tyrosine. Phosphatases, in contrast, remove these phosphate groups.
Extracellular signals including hormones, neurotransmitters, and growth and differentiation factors
can activate kinases, which can occur as cell surface receptors or as the activator of the final effector
protein, as well as other locations along the signal transduction pathway. Cascades of kinases occur,
20 as well as kinases sensitive to second messenger molecules. This system allows for the amplification
of weak signals (low abundance growth factor molecules, for example), as well as the synthesis of
many weak signals into an all-or-nothing response. Phosphatases, then, are essential in determining
the extent of phosphorylation in the cell and, together with kinases, regulate key cellular processes
such as metabolic enzyme activity, proliferation, cell growth and differentiation, cell adhesion, and
25 cell cycle progression.

KINASES

Kinases comprise the largest known enzyme superfamily and vary widely in their target
molecules. Kinases catalyze the transfer of high energy phosphate groups from a phosphate donor to
a phosphate acceptor. Nucleotides usually serve as the phosphate donor in these reactions, with most
30 kinases utilizing adenosine triphosphate (ATP). The phosphate acceptor can be any of a variety of
molecules, including nucleosides, nucleotides, lipids, carbohydrates, and proteins. Proteins are
phosphorylated on hydroxyamino acids. Addition of a phosphate group alters the local charge on the
acceptor molecule, causing internal conformational changes and potentially influencing
intermolecular contacts. Reversible protein phosphorylation is the primary method for regulating
35 protein activity in eukaryotic cells. In general, proteins are activated by phosphorylation in response

to extracellular signals such as hormones, neurotransmitters, and growth and differentiation factors. The activated proteins initiate the cell's intracellular response by way of intracellular signaling pathways and second messenger molecules such as cyclic nucleotides, calcium-calmodulin, inositol, and various mitogens, that regulate protein phosphorylation.

5 Kinases are involved in all aspects of a cell's function, from basic metabolic processes, such as glycolysis, to cell-cycle regulation, differentiation, and communication with the extracellular environment through signal transduction cascades. Inappropriate phosphorylation of proteins in cells has been linked to changes in cell cycle progression and cell differentiation. Changes in the cell cycle have been linked to induction of apoptosis or cancer. Changes in cell differentiation have been linked
10 to diseases and disorders of the reproductive system, immune system, and skeletal muscle.

 There are two classes of protein kinases. One class, protein tyrosine kinases (PTKs), phosphorylates tyrosine residues, and the other class, protein serine/threonine kinases (STKs), phosphorylates serine and threonine residues. Some PTKs and STKs possess structural characteristics of both families and have dual specificity for both tyrosine and serine/threonine
15 residues. Almost all kinases contain a conserved 250-300 amino acid catalytic domain containing specific residues and sequence motifs characteristic of the kinase family. The protein kinase catalytic domain can be further divided into 11 subdomains. N-terminal subdomains I-IV fold into a two-lobed structure which binds and orients the ATP donor molecule, and subdomain V spans the two lobes. C-terminal subdomains VI-XI bind the protein substrate and transfer the gamma phosphate from ATP to
20 the hydroxyl group of a tyrosine, serine, or threonine residue. Each of the 11 subdomains contains specific catalytic residues or amino acid motifs characteristic of that subdomain. For example, subdomain I contains an 8-amino acid glycine-rich ATP binding consensus motif, subdomain II contains a critical lysine residue required for maximal catalytic activity, and subdomains VI through IX comprise the highly conserved catalytic core. PTKs and STKs also contain distinct sequence
25 motifs in subdomains VI and VIII which may confer hydroxyamino acid specificity.

 In addition, kinases may also be classified by additional amino acid sequences, generally between 5 and 100 residues, which either flank or occur within the kinase domain. These additional amino acid sequences regulate kinase activity and determine substrate specificity. (Reviewed in Hardie, G. and S. Hanks (1995) The Protein Kinase Facts Book, Vol I, pp. 17-20 Academic Press,
30 San Diego CA.). In particular, two protein kinase signature sequences have been identified in the kinase domain, the first containing an active site lysine residue involved in ATP binding, and the second containing an aspartate residue important for catalytic activity. If a protein analyzed includes the two protein kinase signatures, the probability of that protein being a protein kinase is close to 100% (PROSITE: PDOC00100, November 1995).

35 Protein Tyrosine Kinases

Protein tyrosine kinases (PTKs) may be classified as either transmembrane, receptor PTKs or nontransmembrane, nonreceptor PTK proteins. Transmembrane tyrosine kinases function as receptors for most growth factors. Growth factors bind to the receptor tyrosine kinase (RTK), which causes the receptor to phosphorylate itself (autophosphorylation) and specific intracellular second messenger proteins. Growth factors (GF) that associate with receptor PTKs include epidermal GF, platelet-derived GF, fibroblast GF, hepatocyte GF, insulin and insulin-like GFs, nerve GF, vascular endothelial GF, and macrophage colony stimulating factor.

Nontransmembrane, nonreceptor PTKs lack transmembrane regions and, instead, form signaling complexes with the cytosolic domains of plasma membrane receptors. Receptors that function through non-receptor PTKs include those for cytokines and hormones (growth hormone and prolactin), and antigen-specific receptors on T and B lymphocytes.

Many PTKs were first identified as oncogene products in cancer cells in which PTK activation was no longer subject to normal cellular controls. In fact, about one third of the known oncogenes encode PTKs. Furthermore, cellular transformation (oncogenesis) is often accompanied by increased tyrosine phosphorylation activity (Charbonneau, H. and N.K. Tonks (1992) Annu. Rev. Cell Biol. 8:463-493). Regulation of PTK activity may therefore be an important strategy in controlling some types of cancer.

Protein Serine/Threonine Kinases

Protein serine/threonine kinases (STKs) are nontransmembrane proteins. A subclass of STKs are known as ERKs (extracellular signal regulated kinases) or MAPs (mitogen-activated protein kinases) and are activated after cell stimulation by a variety of hormones and growth factors. Cell stimulation induces a signaling cascade leading to phosphorylation of MEK (MAP/ERK kinase) which, in turn, activates ERK via serine and threonine phosphorylation. A varied number of proteins represent the downstream effectors for the active ERK and implicate it in the control of cell proliferation and differentiation, as well as regulation of the cytoskeleton. Activation of ERK is normally transient, and cells possess dual specificity phosphatases that are responsible for its down-regulation. Also, numerous studies have shown that elevated ERK activity is associated with some cancers. Other STKs include the second messenger dependent protein kinases such as the cyclic-AMP dependent protein kinases (PKA), calcium-calmodulin (CaM) dependent protein kinases, and the mitogen-activated protein kinases (MAP); the cyclin-dependent protein kinases; checkpoint and cell cycle kinases; Numb-associated kinase (Nak); human Fused (hFu); proliferation-related kinases; 5'-AMP-activated protein kinases; and kinases involved in apoptosis.

One member of the ERK family of MAP kinases, ERK 7, is a novel 61-kDa protein that has motif similarities to ERK1 and ERK2, but is not activated by extracellular stimuli as are ERK1 and ERK2 nor by the common activators, c-Jun N-terminal kinase (JNK) and p38 kinase. ERK7 regulates

its nuclear localization and inhibition of growth through its C-terminal tail, not through the kinase domain as is typical with other MAP kinases (Abe, M.K. (1999) Mol. Cell. Biol. 19:1301-1312).

The second messenger dependent protein kinases primarily mediate the effects of second messengers such as cyclic AMP (cAMP), cyclic GMP, inositol triphosphate, phosphatidylinositol, 3,4,5-triphosphate, cyclic ADP ribose, arachidonic acid, diacylglycerol and calcium-calmodulin. The PKAs are involved in mediating hormone-induced cellular responses and are activated by cAMP produced within the cell in response to hormone stimulation. cAMP is an intracellular mediator of hormone action in all animal cells that have been studied. Hormone-induced cellular responses include thyroid hormone secretion, cortisol secretion, progesterone secretion, glycogen breakdown, bone resorption, and regulation of heart rate and force of heart muscle contraction. PKA is found in all animal cells and is thought to account for the effects of cAMP in most of these cells. Altered PKA expression is implicated in a variety of disorders and diseases including cancer, thyroid disorders, diabetes, atherosclerosis, and cardiovascular disease (Isselbacher, K.J. et al. (1994) Harrison's Principles of Internal Medicine, McGraw-Hill, New York NY, pp. 416-431, 1887).

The casein kinase I (CKI) gene family is another subfamily of serine/threonine protein kinases. This continuously expanding group of kinases have been implicated in the regulation of numerous cytoplasmic and nuclear processes, including cell metabolism and DNA replication and repair. CKI enzymes are present in the membranes, nucleus, cytoplasm and cytoskeleton of eukaryotic cells, and on the mitotic spindles of mammalian cells (Fish, K.J. et al. (1995) J. Biol. Chem. 270:14875-14883).

The CKI family members all have a short amino-terminal domain of 9-76 amino acids, a highly conserved kinase domain of 284 amino acids, and a variable carboxyl-terminal domain that ranges from 24 to over 200 amino acids in length (Cegielska, A. et al. (1998) J. Biol. Chem. 273:1357-1364). The CKI family is comprised of highly related proteins, as seen by the identification of isoforms of casein kinase I from a variety of sources. There are at least five mammalian isoforms, α , β , γ , δ , and ϵ . Fish et al. identified CKI-epsilon from a human placenta cDNA library. It is a basic protein of 416 amino acids and is closest to CKI-delta. Through recombinant expression, it was determined to phosphorylate known CKI substrates and was inhibited by the CKI-specific inhibitor CKI-7. The human gene for CKI-epsilon was able to rescue yeast with a slow-growth phenotype caused by deletion of the yeast CKI locus, HRR250 (Fish et al., *supra*).

The mammalian circadian mutation tau was found to be a semidominant autosomal allele of CKI-epsilon that markedly shortens period length of circadian rhythms in Syrian hamsters. The tau locus is encoded by casein kinase I-epsilon, which is also a homolog of the *Drosophila* circadian gene double-time. Studies of both the wildtype and tau mutant CKI-epsilon enzyme indicated that the mutant enzyme has a noticeable reduction in the maximum velocity and autophosphorylation state.

Further, *in vitro*, CKI-epsilon is able to interact with mammalian PERIOD proteins, while the mutant enzyme is deficient in its ability to phosphorylate PERIOD. Lowrey et al. have proposed that CKI-epsilon plays a major role in delaying the negative feedback signal within the transcription-translation-based autoregulatory loop that composes the core of the circadian mechanism. Therefore
5 the CKI-epsilon enzyme is an ideal target for pharmaceutical compounds influencing circadian rhythms, jet-lag and sleep, in addition to other physiologic and metabolic processes under circadian regulation (Lowrey, P.L. et al. (2000) Science 288:483-491).

Homeodomain-interacting protein kinases (HIPKs) are serine/threonine kinases and novel members of the DYRK kinase subfamily (Hofmann, T.G. et al. (2000) Biochimie 82:1123-1127).
10 HIPKs contain a conserved protein kinase domain separated from a domain that interacts with homeoproteins. HIPKs are nuclear kinases, and HIPK2 is highly expressed in neuronal tissue (Kim, Y.H. et al. (1998) J. Biol. Chem. 273:25875-25879; Wang, Y. et al. (2001) Biochim. Biophys. Acta 1518:168-172). HIPKs act as corepressors for homeodomain transcription factors. This corepressor activity is seen in posttranslational modifications such as ubiquitination and phosphorylation, each of
15 which are important in the regulation of cellular protein function (Kim, Y.H. et al. (1999) Proc. Natl. Acad. Sci. USA 96:12350-12355).

The human h-warts protein, a homolog of *Drosophila* warts tumor suppressor gene, maps to chromosome 6q24-25.1. It has a serine/threonine kinase domain and is localized to centrosomes in interphase cells. It is involved in mitosis and functions as a component of the mitotic apparatus
20 (Nishiyama, Y. et al. (1999) FEBS Lett. 459:159-165).

Calcium-Calmodulin Dependent Protein Kinases

Calcium-calmodulin dependent (CaM) kinases are involved in regulation of smooth muscle contraction, glycogen breakdown (phosphorylase kinase), and neurotransmission (CaM kinase I and CaM kinase II). CaM dependent protein kinases are activated by calmodulin, an intracellular calcium
25 receptor, in response to the concentration of free calcium in the cell. Many CaM kinases are also activated by phosphorylation. Some CaM kinases are also activated by autophosphorylation or by other regulatory kinases. CaM kinase I phosphorylates a variety of substrates including the neurotransmitter-related proteins synapsin I and II, the gene transcription regulator, CREB, and the cystic fibrosis conductance regulator protein, CFTR (Haribabu, B. et al. (1995) EMBO J. 14:3679-
30 3686). CaM kinase II also phosphorylates synapsin at different sites and controls the synthesis of catecholamines in the brain through phosphorylation and activation of tyrosine hydroxylase. CaM kinase II controls the synthesis of catecholamines and serotonin, through phosphorylation/activation of tyrosine hydroxylase and tryptophan hydroxylase, respectively (Fujisawa, H. (1990) BioEssays 12:27-29). The mRNA encoding a calmodulin-binding protein kinase-like protein was found to be
35 enriched in mammalian forebrain. This protein is associated with vesicles in both axons and

dendrites and accumulates largely postnatally. The amino acid sequence of this protein is similar to CaM-dependent STKs, and the protein binds calmodulin in the presence of calcium (Godbout, M. et al. (1994) J. Neurosci. 14:1-13).

Mitogen-Activated Protein Kinases

5 The mitogen-activated protein kinases (MAP), which mediate signal transduction from the cell surface to the nucleus via phosphorylation cascades, are another STK family that regulates intracellular signaling pathways. Several subgroups have been identified, and each manifests different substrate specificities and responds to distinct extracellular stimuli (Egan, S.E. and R.A. Weinberg (1993) Nature 365:781-783). There are three kinase modules comprising the MAP kinase
10 cascade: MAPK (MAP), MAPK kinase (MAP2K, MAPKK, or MKK), and MKK kinase (MAP3K, MAPKKK, OR MEKK) (Wang, X.S. et al. (1998) Biochem. Biophys. Res. Commun. 253:33-37). The extracellular-regulated kinase (ERK) pathway is activated by growth factors and mitogens, for example, epidermal growth factor (EGF), ultraviolet light, hyperosmolar medium, heat shock, or endotoxic lipopolysaccharide (LPS). The closely related though distinct parallel pathways, the c-Jun
15 N-terminal kinase (JNK), or stress-activated kinase (SAPK) pathway, and the p38 kinase pathway are activated by stress stimuli and proinflammatory cytokines such as tumor necrosis factor (TNF) and interleukin-1 (IL-1). Altered MAP kinase expression is implicated in a variety of disease conditions including cancer, inflammation, immune disorders, and disorders affecting growth and development.. MAP kinase signaling pathways are present in mammalian cells as well as in yeast.

Cyclin-Dependent Protein Kinases

 The cyclin-dependent protein kinases (CDKs) are STKs that control the progression of cells through the cell cycle. The entry and exit of a cell from mitosis are regulated by the synthesis and destruction of a family of activating proteins called cyclins. Cyclins are small regulatory proteins that bind to and activate CDKs, which then phosphorylate and activate selected proteins involved in the
25 mitotic process. CDKs are unique in that they require multiple inputs to become activated. In addition to cyclin binding, CDK activation requires the phosphorylation of a specific threonine residue and the dephosphorylation of a specific tyrosine residue on the CDK.

 Another family of STKs associated with the cell cycle are the NIMA (never in mitosis)-related kinases (Neks). Both CDKs and Neks are involved in duplication, maturation, and separation
30 of the microtubule organizing center, the centrosome, in animal cells (Fry, A.M. et al. (1998) EMBO J. 17:470-481).

Checkpoint and Cell Cycle Kinases

 In the process of cell division, the order and timing of cell cycle transitions are under control of cell cycle checkpoints, which ensure that critical events such as DNA replication and chromosome
35 segregation are carried out with precision. If DNA is damaged, e.g. by radiation, a checkpoint

pathway is activated that arrests the cell cycle to provide time for repair. If the damage is extensive, apoptosis is induced. In the absence of such checkpoints, the damaged DNA is inherited by aberrant cells which may cause proliferative disorders such as cancer. Protein kinases play an important role in this process. For example, a specific kinase, checkpoint kinase 1 (Chk1), has been identified in yeast and mammals, and is activated by DNA damage in yeast. Activation of Chk1 leads to the arrest of the cell at the G2/M transition (Sanchez, Y. et al. (1997) Science 277:1497-1501). Specifically, Chk1 phosphorylates the cell division cycle phosphatase CDC25, inhibiting its normal function which is to dephosphorylate and activate the cyclin-dependent kinase Cdc2. Cdc2 activation controls the entry of cells into mitosis (Peng, C.-Y. et al. (1997) Science 277:1501-1505). Thus, activation of Chk1 prevents the damaged cell from entering mitosis. A deficiency in a checkpoint kinase, such as Chk1, may also contribute to cancer by failure to arrest cells with damaged DNA at other checkpoints such as G2/M.

Proliferation-Related Kinases

Proliferation-related kinase is a serum/cytokine inducible STK that is involved in regulation of the cell cycle and cell proliferation in human megakaryocytic cells (Li, B. et al. (1996) J. Biol. Chem. 271:19402-19408). Proliferation-related kinase is related to the polo (derived from *Drosophila* polo gene) family of STKs implicated in cell division. Proliferation-related kinase is downregulated in lung tumor tissue and may be a proto-oncogene whose deregulated expression in normal tissue leads to oncogenic transformation.

5'-AMP-activated protein kinase

A ligand-activated STK protein kinase is 5'-AMP-activated protein kinase (AMPK) (Gao, G. et al. (1996) J. Biol Chem. 271:8675-8681). Mammalian AMPK is a regulator of fatty acid and sterol synthesis through phosphorylation of the enzymes acetyl-CoA carboxylase and hydroxymethylglutaryl-CoA reductase and mediates responses of these pathways to cellular stresses such as heat shock and depletion of glucose and ATP. AMPK is a heterotrimeric complex comprised of a catalytic alpha subunit and two non-catalytic beta and gamma subunits that are believed to regulate the activity of the alpha subunit. Subunits of AMPK have a much wider distribution in non-lipogenic tissues such as brain, heart, spleen, and lung than expected. This distribution suggests that its role may extend beyond regulation of lipid metabolism alone.

Kinases in Apoptosis

Apoptosis is a highly regulated signaling pathway leading to cell death that plays a crucial role in tissue development and homeostasis. Deregulation of this process is associated with the pathogenesis of a number of diseases including autoimmune diseases, neurodegenerative disorders, and cancer. Various STKs play key roles in this process. ZIP kinase is an STK containing a C-terminal leucine zipper domain in addition to its N-terminal protein kinase domain. This

C-terminal domain appears to mediate homodimerization and activation of the kinase as well as interactions with transcription factors such as activating transcription factor, ATF4, a member of the cyclic-AMP responsive element binding protein (ATF/CREB) family of transcriptional factors (Sanjo, H. et al. (1998) J. Biol. Chem. 273:29066-29071). DRAK1 and DRAK2 are STKs that share
5 homology with the death-associated protein kinases (DAP kinases), known to function in interferon- γ induced apoptosis (Sanjo et al., *supra*). Like ZIP kinase, DAP kinases contain a C-terminal protein-protein interaction domain, in the form of ankyrin repeats, in addition to the N-terminal kinase domain. ZIP, DAP, and DRAK kinases induce morphological changes associated with apoptosis when transfected into NIH3T3 cells (Sanjo et al., *supra*). However, deletion of either the
10 N-terminal kinase catalytic domain or the C-terminal domain of these proteins abolishes apoptosis activity, indicating that in addition to the kinase activity, activity in the C-terminal domain is also necessary for apoptosis, possibly as an interacting domain with a regulator or a specific substrate.

RICK is another STK recently identified as mediating a specific apoptotic pathway involving the death receptor, CD95 (Inohara, N. et al. (1998) J. Biol. Chem. 273:12296-12300). CD95 is a
15 member of the tumor necrosis factor receptor superfamily and plays a critical role in the regulation and homeostasis of the immune system (Nagata, S. (1997) Cell 88:355-365). The CD95 receptor signaling pathway involves recruitment of various intracellular molecules to a receptor complex following ligand binding. This process includes recruitment of the cysteine protease caspase-8 which, in turn, activates a caspase cascade leading to cell death. RICK is composed of an N-terminal
20 kinase catalytic domain and a C-terminal "caspase-recruitment" domain that interacts with caspase-like domains, indicating that RICK plays a role in the recruitment of caspase-8. This interpretation is supported by the fact that the expression of RICK in human 293T cells promotes activation of caspase-8 and potentiates the induction of apoptosis by various proteins involved in the CD95 apoptosis pathway (Inohara et al., *supra*).

25 Mitochondrial Protein Kinases

A novel class of eukaryotic kinases, related by sequence to prokaryotic histidine protein kinases, are the mitochondrial protein kinases (MPKs) which seem to have no sequence similarity with other eukaryotic protein kinases. These protein kinases are located exclusively in the mitochondrial matrix space and may have evolved from genes originally present in respiration-
30 dependent bacteria which were endocytosed by primitive eukaryotic cells. MPKs are responsible for phosphorylation and inactivation of the branched-chain alpha-ketoacid dehydrogenase and pyruvate dehydrogenase complexes (Harris, R.A. et al. (1995) Adv. Enzyme Regul. 34:147-162). Five MPKs have been identified. Four members correspond to pyruvate dehydrogenase kinase isozymes, regulating the activity of the pyruvate dehydrogenase complex, which is an important regulatory
35 enzyme at the interface between glycolysis and the citric acid cycle. The fifth member corresponds to

a branched-chain alpha-ketoacid dehydrogenase kinase, important in the regulation of the pathway for the disposal of branched-chain amino acids. (Harris, R.A. et al. (1997) Adv. Enzyme Regul. 37:271-293). Both starvation and the diabetic state are known to result in a great increase in the activity of the pyruvate dehydrogenase kinase in the liver, heart and muscle of the rat. This increase contributes
 5 in both disease states to the phosphorylation and inactivation of the pyruvate dehydrogenase complex and conservation of pyruvate and lactate for gluconeogenesis (Harris (1995), *supra*).

KINASES WITH NON-PROTEIN SUBSTRATES

Lipid and Inositol kinases

10 Lipid kinases phosphorylate hydroxyl residues on lipid head groups. A family of kinases involved in phosphorylation of phosphatidylinositol (PI) has been described, each member phosphorylating a specific carbon on the inositol ring (Leervers, S.J. et al. (1999) Curr. Opin. Cell. Biol. 11:219-225). The phosphorylation of phosphatidylinositol is involved in activation of the protein kinase C signaling pathway. The inositol phospholipids (phosphoinositides) intracellular
 15 signaling pathway begins with binding of a signaling molecule to a G-protein linked receptor in the plasma membrane. This leads to the phosphorylation of phosphatidylinositol (PI) residues on the inner side of the plasma membrane by inositol kinases, thus converting PI residues to the biphosphate state (PIP₂). PIP₂ is then cleaved into inositol triphosphate (IP₃) and diacylglycerol. These two products act as mediators for separate signaling pathways. Cellular responses that are mediated by
 20 these pathways are glycogen breakdown in the liver in response to vasopressin, smooth muscle contraction in response to acetylcholine, and thrombin-induced platelet aggregation.

PI 3-kinase (PI3K), which phosphorylates the D3 position of PI and its derivatives, has a central role in growth factor signal cascades involved in cell growth, differentiation, and metabolism. PI3K is a heterodimer consisting of an adapter subunit and a catalytic subunit. The adapter subunit
 25 acts as a scaffolding protein, interacting with specific tyrosine-phosphorylated proteins, lipid moieties, and other cytosolic factors. When the adapter subunit binds tyrosine phosphorylated targets, such as the insulin responsive substrate (IRS)-1, the catalytic subunit is activated and converts PI (4,5) biphosphate (PIP₂) to PI (3,4,5) P₃ (PIP₃). PIP₃ then activates a number of other proteins, including PKA, protein kinase B (PKB), protein kinase C (PKC), glycogen synthase kinase (GSK)-3,
 30 and p70 ribosomal s6 kinase. PI3K also interacts directly with the cytoskeletal organizing proteins, Rac, rho, and cdc42 (Shepherd, P.R. et al. (1998) Biochem. J. 333:471-490). Animal models for diabetes, such as *obese* and *fat* mice, have altered PI3K adapter subunit levels. Specific mutations in the adapter subunit have also been found in an insulin-resistant Danish population, suggesting a role for PI3K in type-2 diabetes (Shepard, *supra*).

35 An example of lipid kinase phosphorylation activity is the phosphorylation of

D-erythro-sphingosine to the sphingolipid metabolite, sphingosine-1-phosphate (SPP). SPP has emerged as a novel lipid second-messenger with both extracellular and intracellular actions (Kohama, T. et al. (1998) J. Biol. Chem. 273:23722-23728). Extracellularly, SPP is a ligand for the G-protein coupled receptor EDG-1 (endothelial-derived, G-protein coupled receptor). Intracellularly, SPP
5 regulates cell growth, survival, motility, and cytoskeletal changes. SPP levels are regulated by sphingosine kinases that specifically phosphorylate D-erythro-sphingosine to SPP. The importance of sphingosine kinase in cell signaling is indicated by the fact that various stimuli, including platelet-derived growth factor (PDGF), nerve growth factor, and activation of protein kinase C, increase cellular levels of SPP by activation of sphingosine kinase, and the fact that competitive
10 inhibitors of the enzyme selectively inhibit cell proliferation induced by PDGF (Kohama et al., *supra*).

Purine Nucleotide Kinases

The purine nucleotide kinases, adenylate kinase (ATP:AMP phosphotransferase, or AdK) and guanylate kinase (ATP:GMP phosphotransferase, or GuK) play a key role in nucleotide metabolism
15 and are crucial to the synthesis and regulation of cellular levels of ATP and GTP, respectively. These two molecules are precursors in DNA and RNA synthesis in growing cells and provide the primary source of biochemical energy in cells (ATP), and signal transduction pathways (GTP). Inhibition of various steps in the synthesis of these two molecules has been the basis of many antiproliferative drugs for cancer and antiviral therapy (Pillwein, K. et al. (1990) Cancer Res. 50:1576-1579).

AdK is found in almost all cell types and is especially abundant in cells having high rates of ATP synthesis and utilization such as skeletal muscle. In these cells AdK is physically associated with mitochondria and myofibrils, the subcellular structures that are involved in energy production and utilization, respectively. Recent studies have demonstrated a major function for AdK in
20 transferring high energy phosphoryls from metabolic processes generating ATP to cellular components consuming ATP (Zelevnikar, R.J. et al. (1995) J. Biol. Chem. 270:7311-7319). Thus AdK may have a pivotal role in maintaining energy production in cells, particularly those having a high rate of growth or metabolism such as cancer cells, and may provide a target for suppression of its activity in order to treat certain cancers. Alternatively, reduced AdK activity may be a source of various metabolic, muscle-energy disorders that can result in cardiac or respiratory failure and may be
25 treatable by increasing AdK activity.

GuK, in addition to providing a key step in the synthesis of GTP for RNA and DNA synthesis, also fulfills an essential function in signal transduction pathways of cells through the regulation of GDP and GTP. Specifically, GTP binding to membrane associated G proteins mediates the activation of cell receptors, subsequent intracellular activation of adenyl cyclase, and production
35 of the second messenger, cyclic AMP. GDP binding to G proteins inhibits these processes. GDP and

GTP levels also control the activity of certain oncogenic proteins such as p21^{ras} known to be involved in control of cell proliferation and oncogenesis (Bos, J.L. (1989) Cancer Res. 49:4682-4689). High ratios of GTP:GDP caused by suppression of GuK cause activation of p21^{ras} and promote oncogenesis. Increasing GuK activity to increase levels of GDP and reduce the GTP:GDP ratio may provide a therapeutic strategy to reverse oncogenesis.

GuK is an important enzyme in the phosphorylation and activation of certain antiviral drugs useful in the treatment of herpes virus infections. These drugs include the guanine homologs acyclovir and buciclovir (Miller, W.H. and R.L. Miller (1980) J. Biol. Chem. 255:7204-7207; Stenberg, K. et al. (1986) J. Biol. Chem. 261:2134-2139). Increasing GuK activity in infected cells may provide a therapeutic strategy for augmenting the effectiveness of these drugs and possibly for reducing the necessary dosages of the drugs.

Pyrimidine Kinases

The pyrimidine kinases are deoxycytidine kinase and thymidine kinase 1 and 2. Deoxycytidine kinase is located in the nucleus, and thymidine kinase 1 and 2 are found in the cytosol (Johansson, M. et al. (1997) Proc. Natl. Acad. Sci. USA 94:11941-11945). Phosphorylation of deoxyribonucleosides by pyrimidine kinases provides an alternative pathway for *de novo* synthesis of DNA precursors. The role of pyrimidine kinases, like purine kinases, in phosphorylation is critical to the activation of several chemotherapeutically important nucleoside analogues (Arner E.S. and S. Eriksson (1995) Pharmacol. Ther. 67:155-186).

Pantothenate Kinases

Pantothenate kinase (PanK) is a key regulatory enzyme in the CoA biosynthetic pathway in bacteria. It catalyzes the phosphorylation of pantothenic acid to form phosphopantothenate. CoA is a feedback inhibitor of PanK activity by competitive binding to the ATP site. Even though the predicted protein sequence of mammalian (PanK) is not related to bacterial PanK, it too is a key regulatory enzyme in mammalian CoA biosynthesis (Rock, C.O. (2000) J. Biol. Chem. 275:1377-1383). PanK is regulated by feedback inhibition of CoA and its acyl esters. This inhibition is modified by changes in the concentration of free carnitine, a nonessential activator of PanK (Fisher, M.N. et al. (1985) J. Biol. Chem. 260:15745-15751).

PHOSPHATASES

Protein phosphatases are generally characterized as either serine/threonine- or tyrosine-specific based on their preferred phospho-amino acid substrate. However, some phosphatases (DSPs, for dual specificity phosphatases) can act on phosphorylated tyrosine, serine, or threonine residues. The protein serine/threonine phosphatases (PSPs) are important regulators of many cAMP-mediated hormone responses in cells. Protein tyrosine phosphatases (PTPs) play a significant role in cell cycle

and cell signaling processes. Another family of phosphatases is the acid phosphatase or histidine acid phosphatase (HAP) family whose members hydrolyze phosphate esters at acidic pH conditions.

PSPs are found in the cytosol, nucleus, and mitochondria and in association with cytoskeletal and membranous structures in most tissues, especially the brain. Some PSPs require divalent cations, such as Ca^{2+} or Mn^{2+} , for activity. PSPs play important roles in glycogen metabolism, muscle contraction, protein synthesis, T cell function, neuronal activity, oocyte maturation, and hepatic metabolism (reviewed in Cohen, P. (1989) *Annu. Rev. Biochem.* 58:453-508). PSPs can be separated into two classes. The PPP class includes PP1, PP2A, PP2B/calcineurin, PP4, PP5, PP6, and PP7. Members of this class are composed of a homologous catalytic subunit bearing a very highly conserved signature sequence, coupled with one or more regulatory subunits (PROSITE PDOC00115). Further interactions with scaffold and anchoring molecules determine the intracellular localization of PSPs and substrate specificity. The PPM class consists of several closely related isoforms of PP2C and is evolutionarily unrelated to the PPP class.

PP1 dephosphorylates many of the proteins phosphorylated by cyclic AMP-dependent protein kinase (PKA) and is an important regulator of many cAMP-mediated hormone responses in cells. A number of isoforms have been identified, with the alpha and beta forms being produced by alternative splicing of the same gene. Both ubiquitous and tissue-specific targeting proteins for PP1 have been identified. In the brain, inhibition of PP1 activity by the dopamine and adenosine 3',5'-monophosphate-regulated phosphoprotein of 32kDa (DARPP-32) is necessary for normal dopamine response in neostriatal neurons (reviewed in Price, N.E. and M.C. Mumby (1999) *Curr. Opin. Neurobiol.* 9:336-342). PP1, along with PP2A, has been shown to limit motility in microvascular endothelial cells, suggesting a role for PSPs in the inhibition of angiogenesis (Gabel, S. et al. (1999) *Otolaryngol. Head Neck Surg.* 121:463-468).

PP2A is the main serine/threonine phosphatase. The core PP2A enzyme consists of a single 36 kDa catalytic subunit (C) associated with a 65 kDa scaffold subunit (A), whose role is to recruit additional regulatory subunits (B). Three gene families encoding B subunits are known (PR55, PR61, and PR72), each of which contain multiple isoforms, and additional families may exist (Millward, T.A et al. (1999) *Trends Biosci.* 24:186-191). These "B-type" subunits are cell type- and tissue-specific and determine the substrate specificity, enzymatic activity, and subcellular localization of the holoenzyme. The PR55 family is highly conserved and bears a conserved motif (PROSITE PDOC00785). PR55 increases PP2A activity toward mitogen-activated protein kinase (MAPK) and MAPK kinase (MEK). PP2A dephosphorylates the MAPK active site, inhibiting the cell's entry into mitosis. Several proteins can compete with PR55 for PP2A core enzyme binding, including the CKII kinase catalytic subunit, polyomavirus middle and small T antigens, and SV40 small t antigen. Viruses may use this mechanism to commandeer PP2A and stimulate progression of the cell through

the cell cycle (Pallas, D.C. et al. (1992) J. Virol. 66:886-893). Altered MAP kinase expression is also implicated in a variety of disease conditions including cancer, inflammation, immune disorders, and disorders affecting growth and development. PP2A, in fact, can dephosphorylate and modulate the activities of more than 30 protein kinases *in vitro*, and other evidence suggests that the same is true *in vivo* for such kinases as PKB, PKC, the calmodulin-dependent kinases, ERK family MAP kinases, cyclin-dependent kinases, and the I κ B kinases (reviewed in Millward et al., *supra*). PP2A is itself a substrate for CKI and CKII kinases, and can be stimulated by polycationic macromolecules. A PP2A-like phosphatase is necessary to maintain the G1 phase destruction of mammalian cyclins A and B (Bastians, H. et al. (1999) Mol. Biol. Cell 10:3927-3941). PP2A is a major activity in the brain and is implicated in regulating neurofilament stability and normal neural function, particularly the phosphorylation of the microtubule-associated protein tau. Hyperphosphorylation of tau has been proposed to lead to the neuronal degeneration seen in Alzheimer's disease (reviewed in Price and Mumby, *supra*).

PP2B, or calcineurin, is a Ca²⁺-activated dimeric phosphatase and is particularly abundant in the brain. It consists of catalytic and regulatory subunits, and is activated by the binding of the calcium/calmodulin complex. Calcineurin is the target of the immunosuppressant drugs cyclosporine and FK506. Along with other cellular factors, these drugs interact with calcineurin and inhibit phosphatase activity. In T cells, this blocks the calcium dependent activation of the NF-AT family of transcription factors, leading to immunosuppression. This family is widely distributed, and it is likely that calcineurin regulates gene expression in other tissues as well. In neurons, calcineurin modulates functions which range from the inhibition of neurotransmitter release to desensitization of postsynaptic NMDA-receptor coupled calcium channels to long term memory (reviewed in Price and Mumby, *supra*).

Other members of the PPP class have recently been identified (Cohen, P.T. (1997) Trends Biochem. Sci. 22:245-251). One of them, PP5, contains regulatory domains with tetratricopeptide repeats. It can be activated by polyunsaturated fatty acids and anionic phospholipids *in vitro* and appears to be involved in a number of signaling pathways, including those controlled by atrial natriuretic peptide or steroid hormones (reviewed in Andreeva, A.V. and M.A. Kutuzov (1999) Cell Signal. 11:555-562).

PP2C is a ~42kDa monomer with broad substrate specificity and is dependent on divalent cations (mainly Mn²⁺ or Mg²⁺) for its activity. PP2C proteins share a conserved N-terminal region with an invariant DGH motif, which contains an aspartate residue involved in cation binding (PROSITE PDOC00792). Targeting proteins and mechanisms regulating PP2C activity have not been identified. PP2C has been shown to inhibit the stress-responsive p38 and Jun kinase (JNK) pathways (Takekawa, M. et al. (1998) EMBO J. 17:4744-4752).

In contrast to PSPs, tyrosine-specific phosphatases (PTPs) are generally monomeric proteins of very diverse size (from 20kDa to greater than 100kDa) and structure that function primarily in the transduction of signals across the plasma membrane. PTPs are categorized as either soluble phosphatases or transmembrane receptor proteins that contain a phosphatase domain. All PTPs share a conserved catalytic domain of about 300 amino acids which contains the active site. The active site consensus sequence includes a cysteine residue which executes a nucleophilic attack on the phosphate moiety during catalysis (Neel, B.G. and N.K. Tonks (1997) *Curr. Opin. Cell Biol.* 9:193-204). Receptor PTPs are made up of an N-terminal extracellular domain of variable length, a transmembrane region, and a cytoplasmic region that generally contains two copies of the catalytic domain. Although only the first copy seems to have enzymatic activity, the second copy apparently affects the substrate specificity of the first. The extracellular domains of some receptor PTPs contain fibronectin-like repeats, immunoglobulin-like domains, MAM domains (an extracellular motif likely to have an adhesive function), or carbonic anhydrase-like domains (PROSITE PDOC 00323). This wide variety of structural motifs accounts for the diversity in size and specificity of PTPs.

PTPs play important roles in biological processes such as cell adhesion, lymphocyte activation, and cell proliferation. PTPs μ and κ are involved in cell-cell contacts, perhaps regulating cadherin/catenin function. A number of PTPs affect cell spreading, focal adhesions, and cell motility, most of them via the integrin/tyrosine kinase signaling pathway (reviewed in Neel and Tonks, *supra*). CD45 phosphatases regulate signal transduction and lymphocyte activation (Ledbetter, J.A. et al. (1988) *Proc. Natl. Acad. Sci. USA* 85:8628-8632). Soluble PTPs containing Src-homology-2 domains have been identified (SHPs), suggesting that these molecules might interact with receptor tyrosine kinases. SHP-1 regulates cytokine receptor signaling by controlling the Janus family PTKs in hematopoietic cells, as well as signaling by the T-cell receptor and c-Kit (reviewed in Neel and Tonks, *supra*). M-phase inducer phosphatase plays a key role in the induction of mitosis by dephosphorylating and activating the PTK CDC2, leading to cell division (Sadhu, K. et al. (1990) *Proc. Natl. Acad. Sci. USA* 87:5139-5143). In addition, the genes encoding at least eight PTPs have been mapped to chromosomal regions that are translocated or rearranged in various neoplastic conditions, including lymphoma, small cell lung carcinoma, leukemia, adenocarcinoma, and neuroblastoma (reviewed in Charbonneau, H. and N.K. Tonks (1992) *Annu. Rev. Cell Biol.* 8:463-493). The PTP enzyme active site comprises the consensus sequence of the MTM1 gene family. The MTM1 gene is responsible for X-linked recessive myotubular myopathy, a congenital muscle disorder that has been linked to Xq28 (Kioschis, P. et al., (1998) *Genomics* 54:256-266). Many PTKs are encoded by oncogenes, and it is well known that oncogenesis is often accompanied by increased tyrosine phosphorylation activity. It is therefore possible that PTPs may serve to prevent or reverse cell transformation and the growth of various cancers by controlling the levels of tyrosine

phosphorylation in cells. This is supported by studies showing that overexpression of PTP can suppress transformation in cells and that specific inhibition of PTP can enhance cell transformation (Charbonneau and Tonks, *supra*).

Dual specificity phosphatases (DSPs) are structurally more similar to the PTPs than the PSPs. DSPs bear an extended PTP active site motif with an additional 7 amino acid residues. DSPs are primarily associated with cell proliferation and include the cell cycle regulators cdc25A, B, and C. The phosphatases DUSP1 and DUSP2 inactivate the MAPK family members ERK (extracellular signal-regulated kinase), JNK (c-Jun N-terminal kinase), and p38 on both tyrosine and threonine residues (PROSITE PDOC 00323, *supra*). In the activated state, these kinases have been implicated in neuronal differentiation, proliferation, oncogenic transformation, platelet aggregation, and apoptosis. Thus, DSPs are necessary for proper regulation of these processes (Muda, M. et al. (1996) J. Biol. Chem. 271:27205-27208). The tumor suppressor PTEN is a DSP that also shows lipid phosphatase activity. It seems to negatively regulate interactions with the extracellular matrix and maintains sensitivity to apoptosis. PTEN has been implicated in the prevention of angiogenesis (Giri, D. and M. Ittmann (1999) Hum. Pathol. 30:419-424) and abnormalities in its expression are associated with numerous cancers (reviewed in Tamura, M. et al. (1999) J. Natl. Cancer Inst. 91:1820-1828).

Histidine acid phosphatase (HAP; EXPASY EC 3.1.3.2), also known as acid phosphatase, hydrolyzes a wide spectrum of substrates including alkyl, aryl, and acyl orthophosphate monoesters and phosphorylated proteins at low pH. HAPs share two regions of conserved sequences, each centered around a histidine residue which is involved in catalytic activity. Members of the HAP family include lysosomal acid phosphatase (LAP) and prostatic acid phosphatase (PAP), both sensitive to inhibition by L-tartrate (PROSITE PDOC00538).

Synaptojanin, a polyphosphoinositide phosphatase, dephosphorylates phosphoinositides at positions 3, 4 and 5 of the inositol ring. Synaptojanin is a major presynaptic protein found at clathrin-coated endocytic intermediates in nerve terminals, and binds the clathrin coat-associated protein, EPS15. This binding is mediated by the C-terminal region of synaptojanin-170, which has 3 Asp-Pro-Phe amino acid repeats. Further, this 3 residue repeat had been found to be the binding site for the EH domains of EPS15 (Haffner, C. et al. (1997) FEBS Lett. 419:175-180). Additionally, synaptojanin may potentially regulate interactions of endocytic proteins with the plasma membrane, and be involved in synaptic vesicle recycling (Brodin, L. et al. (2000) Curr. Opin. Neurobiol. 10:312-320). Studies in mice with a targeted disruption in the synaptojanin 1 gene (Synj1) were shown to support coat formation of endocytic vesicles more effectively than was seen in wild-type mice, suggesting that Synj1 can act as a negative regulator of membrane-coat protein interactions. These findings provide genetic evidence for a crucial role of phosphoinositide metabolism in synaptic

vesicle recycling (Cremona, O. et al. (1999) Cell 99:179-188).

Expression profiling

Array technology can provide a simple way to explore the expression of a single polymorphic gene or the expression profile of a large number of related or unrelated genes. When the expression of a single gene is examined, arrays are employed to detect the expression of a specific gene or its variants. When an expression profile is examined, arrays provide a platform for identifying genes that are tissue specific, are affected by a substance being tested in a toxicology assay, are part of a signaling cascade, carry out housekeeping functions, or are specifically related to a particular genetic predisposition, condition, disease, or disorder.

The discovery of new kinases and phosphatases, and the polynucleotides encoding them, satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cardiovascular diseases, immune system disorders, neurological disorders, disorders affecting growth and development, lipid disorders, cell proliferative disorders, and cancers, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of kinases and phosphatases.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, kinases and phosphatases, referred to collectively as "KPP" and individually as "KPP-1," "KPP-2," "KPP-3," "KPP-4," "KPP-5," "KPP-6," "KPP-7," "KPP-8," "KPP-9," "KPP-10," "KPP-11," "KPP-12," "KPP-13," and "KPP-14." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-14.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14.

In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-14. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:15-28.

5 Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group
10 consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group
15 consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of a polypeptide having an amino
20 acid sequence selected from the group consisting of SEQ ID NO:1-14. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

25 Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an
30 immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at
35 least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID

NO:15-28, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and a pharmaceutically acceptable excipient. In one embodiment, the

composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-14. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional KPP, comprising administering to a patient in need of such treatment the composition.

5 The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, c) a biologically active fragment of a polypeptide
10 having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a
15 pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional KPP, comprising administering to a patient in need of such treatment the composition.

 Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an
20 amino acid sequence selected from the group consisting of SEQ ID NO:1-14, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the
25 group consisting of SEQ ID NO:1-14. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional KPP, comprising
30 administering to a patient in need of such treatment the composition.

 The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected
35 from the group consisting of SEQ ID NO:1-14, c) a biologically active fragment of a polypeptide

having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, b) detecting altered expression of the target polynucleotide, and c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide

in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28, 5 iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in 10 an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide 15 sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog, and the PROTEOME database identification numbers and annotations of PROTEOME database homologs, for polypeptides of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are also shown.

20 Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide 25 sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

30 Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

Table 8 shows single nucleotide polymorphisms found in polynucleotide sequences of the invention, along with allele frequencies in different human populations.

35

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which
5 will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms “a,” “an,” and “the” include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to “a host cell” includes a plurality of such host cells, and a reference to “an antibody” is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so
10 forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now
15 described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

20 “KPP” refers to the amino acid sequences of substantially purified KPP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term “agonist” refers to a molecule which intensifies or mimics the biological activity of KPP. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other
25 compound or composition which modulates the activity of KPP either by directly interacting with KPP or by acting on components of the biological pathway in which KPP participates.

An “allelic variant” is an alternative form of the gene encoding KPP. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or
30 many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

“Altered” nucleic acid sequences encoding KPP include those sequences with deletions,
35 insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as KPP or a

polypeptide with at least one functional characteristic of KPP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding KPP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding KPP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent KPP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of KPP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of KPP. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of KPP either by directly interacting with KPP or by acting on components of the biological pathway in which KPP participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')₂, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind KPP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term “antigenic determinant” refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term “aptamer” refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an *in vitro* evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH₂), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker. (See, e.g., Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13.)

The term “intramer” refers to an aptamer which is expressed *in vivo*. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl. Acad. Sci. USA 96:3606-3610).

The term “spiegelmer” refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

The term “antisense” refers to any composition capable of base-pairing with the “sense” (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation “negative” or “minus” can refer to the antisense strand, and the

designation “positive” or “plus” can refer to the sense strand of a reference DNA molecule.

The term “biologically active” refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, “immunologically active” or “immunogenic” refers to the capability of the natural, recombinant, or synthetic KPP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

“Complementary” describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A “composition comprising a given polynucleotide sequence” and a “composition comprising a given amino acid sequence” refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding KPP or fragments of KPP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

“Consensus sequence” refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

“Conservative amino acid substitutions” are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

	Original Residue	Conservative Substitution
	Ala	Gly, Ser
	Arg	His, Lys
	Asn	Asp, Gln, His
	Asp	Asn, Glu
	Cys	Ala, Ser
	Gln	Asn, Glu, His
	Glu	Asp, Gln, His

	Gly	Ala
	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
	Leu	Ile, Val
5	Lys	Arg, Gln, Glu
	Met	Leu, Ile
	Phe	His, Met, Leu, Trp, Tyr
	Ser	Cys, Thr
	Thr	Ser, Val
10	Trp	Phe, Tyr
	Tyr	His, Phe, Trp
	Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A “deletion” refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term “derivative” refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A “detectable label” refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

“Differential expression” refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

“Exon shuffling” refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A “fragment” is a unique portion of KPP or the polynucleotide encoding KPP which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment

used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected
5 from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:15-28 comprises a region of unique polynucleotide sequence that
10 specifically identifies SEQ ID NO:15-28, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:15-28 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:15-28 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:15-28 and the region of SEQ ID NO:15-28 to which the fragment corresponds are routinely
15 determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-14 is encoded by a fragment of SEQ ID NO:15-28. A fragment of SEQ ID NO:1-14 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-14. For example, a fragment of SEQ ID NO:1-14 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-14. The precise length of
20 a fragment of SEQ ID NO:1-14 and the region of SEQ ID NO:1-14 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A “full length” polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A
25 “full length” polynucleotide sequence encodes a “full length” polypeptide sequence.

“Homology” refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms “percent identity” and “% identity,” as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a
30 standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e
35 sequence alignment program. This program is part of the LASERGENE software package, a suite of

molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at <http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

20 *Matrix: BLOSUM62*
 Reward for match: 1
 Penalty for mismatch: -2
 Open Gap: 5 and Extension Gap: 2 penalties
 Gap x drop-off: 50
 25 *Expect: 10*
 Word Size: 11
 Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

35 Nucleic acid sequences that do not show a high degree of identity may nevertheless encode

similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases “percent identity” and “% identity,” as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and “diagonals saved”=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the “percent similarity” between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the “BLAST 2 Sequences” tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

“Human artificial chromosomes” (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for

chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

5 "Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the
10 stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity.
15 Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

 Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic
20 strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

25 High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents
30 include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such
35 similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of KPP which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of KPP which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of KPP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of KPP.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of

amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

“Post-translational modification” of an KPP may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of KPP.

“Probe” refers to nucleic acid sequences encoding KPP, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes.

“Primers” are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of

Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a “mispriming library,” in which
5 sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user’s specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments,
10 thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to
15 identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A “recombinant nucleic acid” is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the
20 artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, *supra*. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to
25 transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A “regulatory element” refers to a nucleic acid sequence usually derived from untranslated
30 regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

“Reporter molecules” are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent,
35 chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and

other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose
5 instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing KPP, nucleic acids encoding KPP, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

10 The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide
15 comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which
20 they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers,
25 microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" or "expression profile" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient
30 cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term
35 "transformed cells" includes stably transformed cells in which the inserted DNA is capable of

replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. In one alternative, the nucleic acid can be introduced by infection with a recombinant viral vector, such as a lentiviral vector (Lois, C. et al. (2002) Science 295:868-872). The term genetic manipulation does not include classical cross-breeding, or *in vitro* fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), *supra*.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having

at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human kinases and phosphatases (KPP), the polynucleotides encoding KPP, and the use of these compositions for the diagnosis, treatment, or prevention of cardiovascular diseases, immune system disorders, neurological disorders, disorders affecting growth and development, lipid disorders, cell proliferative disorders, and cancers.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown. Column 6 shows the Incyte ID numbers of physical, full length clones corresponding to the polypeptide and polynucleotide sequences of the invention. The full length clones encode polypeptides which have at least 95% sequence identity to the polypeptide sequences shown in column 3.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database and the PROTEOME database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID NO:) of the nearest GenBank homolog and the PROTEOME database identification numbers (PROTEOME ID NO:) of the nearest PROTEOME database homologs. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank and PROTEOME database homolog(s) along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention.

Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are kinases and phosphatases. For example, SEQ ID NO:2 is 93% identical, from residue M1 to residue V1037, to *Mus musculus* serine/threonine kinase UNC51.2 (GenBank ID g6580857) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:2 also contains a protein kinase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, PROFILESCAN, and additional BLAST analyses provide further corroborative evidence that SEQ ID NO:2 is a protein kinase.

In an alternative example, SEQ ID NO:6 is 83% identical, from residue A13 to residue N416, to rat inositol polyphosphate multikinase (GenBank ID g13162658) as determined by BLAST. (See Table 2.) The BLAST probability score is $2.2e-183$.

In an alternative example, SEQ ID NO:7 is 40% identical, from residue Q10 to residue K319, to *Schizosaccharomyces pombe* putative Trp-Asp repeat protein (GenBank ID g 3947883) as determined by BLAST. (See Table 2.) The BLAST probability score is $4.7e-62$. SEQ ID NO:7 also contains a WD domain, G-beta repeat as determined by searching for statistically significant matches in the HMM-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:7 is a Trp-Asp (WD) repeat protein.

In an alternative example, SEQ ID NO:8 is 67% identical, from residue P86 to residue E529 to rat protein kinase WNK1, a novel mammalian serine/threonine protein kinase lacking the catalytic lysine in subdomain II (GenBank ID g8272557) as determined by BLAST. (See Table 2.) The BLAST probability score is $2.6e-200$. SEQ ID NO:8 also contains a protein kinase domain as determined by searching for statistically significant matches in the HMM-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, PROFILESCAN, and other BLAST analyses provide further corroborative evidence that SEQ ID NO:8 is a serine/threonine protein kinase.

In an alternative example, SEQ ID NO:10 is 97% identical, from residue R127 to residue

V548, to the zeta isozyme of human protein kinase C (GenBank ID g35501) as determined by BLAST. (See Table 2.) The BLAST probability score is $3.3e-296$. SEQ ID NO:10 also contains a protein kinase domain and a protein kinase C terminal domain as determined by searching for statistically significant matches in the HMM-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and other BLAST analyses provide further corroborative evidence that SEQ ID NO:10 is a protein kinase C.

SEQ ID NO:1, SEQ ID NO:3-5, SEQ ID NO:9 and SEQ ID NO:11-14 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-14 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Column 1 lists the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:), the corresponding Incyte polynucleotide consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 shows the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences used to assemble the full length polynucleotide sequences of the invention, and of fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:15-28 or that distinguish between SEQ ID NO:15-28 and related polynucleotide sequences.

The polynucleotide fragments described in Column 2 of Table 4 may refer specifically, for example, to Incyte cDNAs derived from tissue-specific cDNA libraries or from pooled cDNA libraries. Alternatively, the polynucleotide fragments described in column 2 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotide sequences. In addition, the polynucleotide fragments described in column 2 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the polynucleotide fragments described in column 2 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the polynucleotide fragments described in column 2 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, a polynucleotide sequence identified as FL_XXXXXX_{N₁}_{N₂}YYYYY_{N₃}_{N₄} represents a "stitched" sequence in which XXXXXX is the identification number of the cluster of sequences to which the algorithm was applied, and YYYYYY is the number of the prediction generated by the algorithm, and $N_{1,2,3...}$, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the

polynucleotide fragments in column 2 may refer to assemblages of exons brought together by an “exon-stretching” algorithm. For example, a polynucleotide sequence identified as FLXXXXXX_gAAAAA_gBBBBB_1_N is a “stretched” sequence, with XXXXXX being the Incyte project identification number, gAAAAA being the GenBank identification number of the human genomic sequence to which the “exon-stretching” algorithm was applied, gBBBBB being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and N referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the “exon-stretching” algorithm, a RefSeq identifier (denoted by “NM,” “NP,” or “NT”) may be used in place of the GenBank identifier (*i.e.*, gBBBBB).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in Table 4 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

Table 8 shows single nucleotide polymorphisms (SNPs) found in polynucleotide sequences of the invention, along with allele frequencies in different human populations. Columns 1 and 2 show the polynucleotide sequence identification number (SEQ ID NO:) and the corresponding Incyte

project identification number (PID) for polynucleotides of the invention. Column 3 shows the Incyte identification number for the EST in which the SNP was detected (EST ID), and column 4 shows the identification number for the SNP (SNP ID). Column 5 shows the position within the EST sequence at which the SNP is located (EST SNP), and column 6 shows the position of the SNP within the full-length polynucleotide sequence (CB1 SNP). Column 7 shows the allele found in the EST sequence. Columns 8 and 9 show the two alleles found at the SNP site. Column 10 shows the amino acid encoded by the codon including the SNP site, based upon the allele found in the EST. Columns 11-14 show the frequency of allele 1 in four different human populations. An entry of n/d (not detected) indicates that the frequency of allele 1 in the population was too low to be detected, while n/a (not available) indicates that the allele frequency was not determined for the population.

The invention also encompasses KPP variants. A preferred KPP variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the KPP amino acid sequence, and which contains at least one functional or structural characteristic of KPP.

The invention also encompasses polynucleotides which encode KPP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:15-28, which encodes KPP. The polynucleotide sequences of SEQ ID NO:15-28, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding KPP. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding KPP. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:15-28 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:15-28. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of KPP.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a polynucleotide sequence encoding KPP. A splice variant may have portions which have significant sequence identity to the polynucleotide sequence encoding KPP, but will generally have a greater or lesser number of polynucleotides due to additions or deletions of blocks of sequence arising from alternate splicing of exons during mRNA processing. A splice variant may have less than about 70%, or alternatively less than about 60%, or alternatively less than about 50% polynucleotide

sequence identity to the polynucleotide sequence encoding KPP over its entire length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide sequence encoding KPP. For example, a polynucleotide comprising a sequence of
5 SEQ ID NO:26 is a splice variant of a polynucleotide comprising a sequence of SEQ ID NO:16, a polynucleotide comprising a sequence of SEQ ID NO:27 is a splice variant of a polynucleotide comprising a sequence of SEQ ID NO:23, and a polynucleotide comprising a sequence of SEQ ID NO:28 is a splice variant of a polynucleotide comprising a sequence of SEQ ID NO:21. Any one of the splice variants described above can encode an amino acid sequence which contains at least one
10 functional or structural characteristic of KPP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding KPP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide
15 sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring KPP, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode KPP and its variants are generally capable of
20 hybridizing to the nucleotide sequence of the naturally occurring KPP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding KPP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which
25 particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding KPP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode KPP and KPP
30 derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding KPP or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of
35 hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID

NO:15-28 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

5 Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Biosciences, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification
10 system (Invitrogen, Carlsbad CA). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Amersham Biosciences), or other
15 systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

 The nucleic acid sequences encoding KPP may be extended utilizing a partial nucleotide
20 sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) *PCR Methods Applic.* 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown
25 sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) *Nucleic Acids Res.* 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) *PCR Methods Applic.* 1:111-119.) In this method, multiple restriction enzyme
30 digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) *Nucleic Acids Res.* 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries
35 and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed

using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

5 When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

10 Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate
15 software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

 In another embodiment of the invention, polynucleotide sequences or fragments thereof
20 which encode KPP may be cloned in recombinant DNA molecules that direct expression of KPP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express KPP.

 The nucleotide sequences of the present invention can be engineered using methods generally
25 known in the art in order to alter KPP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction
30 sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

 The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Cramer, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or
35 improve the biological properties of KPP, such as its biological or enzymatic activity or its ability to

bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding KPP may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) *Nucleic Acids Symp. Ser. 7*:215-223; and Horn, T. et al. (1980) *Nucleic Acids Symp. Ser. 7*:225-232.) Alternatively, KPP itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) *Proteins, Structures and Molecular Properties*, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) *Science* 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of KPP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) *Methods Enzymol.* 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, *supra*, pp. 28-53.)

In order to express a biologically active KPP, the nucleotide sequences encoding KPP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding KPP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding KPP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding KPP and its initiation codon and upstream regulatory sequences are inserted into

the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both
5 natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding KPP and appropriate transcriptional and translational control
10 elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences
15 encoding KPP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or
20 animal cell systems. (See, e.g., Sambrook, *supra*; Ausubel, *supra*; Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509; Engelhard, E.K. et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3224-3227; Sandig, V. et al. (1996) *Hum. Gene Ther.* 7:1937-1945; Takamatsu, N. (1987) *EMBO J.* 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659; and
25 Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) *Cancer Gen. Ther.* 5(6):350-356; Yu, M. et al. (1993) *Proc. Natl. Acad. Sci. USA* 90(13):6340-6344; Buller, R.M. et al. (1985) *Nature* 317(6040):813-815; McGregor, D.P. et al.
30 (1994) *Mol. Immunol.* 31(3):219-226; and Verma, I.M. and N. Somia (1997) *Nature* 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding KPP. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding KPP can be achieved using a
35 multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1

plasmid (Invitrogen). Ligation of sequences encoding KPP into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for *in vitro* transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509.) When large quantities of KPP are needed, e.g. for the production of antibodies, vectors which direct high level expression of KPP may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of KPP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast *Saccharomyces cerevisiae* or *Pichia pastoris*. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, *supra*; Bitter, G.A. et al. (1987) *Methods Enzymol.* 153:516-544; and Scorer, C.A. et al. (1994) *Bio/Technology* 12:181-184.)

Plant systems may also be used for expression of KPP. Transcription of sequences encoding KPP may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) *EMBO J.* 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) *EMBO J.* 3:1671-1680; Broglie, R. et al. (1984) *Science* 224:838-843; and Winter, J. et al. (1991) *Results Probl. Cell Differ.* 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding KPP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses KPP in host cells. (See, e.g., Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino

polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of KPP in cell lines is preferred. For example, sequences encoding KPP can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk* and *apr* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), β glucuronidase and its substrate β -glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding KPP is inserted within a marker gene sequence, transformed cells containing sequences encoding KPP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding KPP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding KPP and that express KPP may be identified by a variety of procedures known to those of skill in the art. These procedures

include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of KPP using either
5 specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on KPP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g.,
10 Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and
15 may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding KPP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding KPP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available,
20 and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Biosciences, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as
25 well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding KPP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors
30 containing polynucleotides which encode KPP may be designed to contain signal sequences which direct secretion of KPP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation,
35 phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or

“pro” form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct
5 modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding KPP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric KPP protein containing a heterologous moiety that can be recognized by a commercially available antibody may
10 facilitate the screening of peptide libraries for inhibitors of KPP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their
15 cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the KPP encoding sequence and the heterologous protein
20 sequence, so that KPP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, *supra*, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled KPP may be achieved *in*
25 *vitro* using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ³⁵S-methionine.

KPP of the present invention or fragments thereof may be used to screen for compounds that
30 specifically bind to KPP. At least one and up to a plurality of test compounds may be screened for specific binding to KPP. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., ligands or receptors), or small molecules. In one embodiment, the compound thus identified is closely related to the natural ligand of KPP, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991)
35 Current Protocols in Immunology 1(2):Chapter 5.) In another embodiment, the compound thus

identified is a natural ligand of a receptor KPP. (See, e.g., Howard, A.D. et al. (2001) Trends Pharmacol. Sci.22:132-140; Wise, A. et al. (2002) Drug Discovery Today 7:235-246.)

In other embodiments, the compound can be closely related to the natural receptor to which KPP binds, at least a fragment of the receptor, or a fragment of the receptor including all or a portion of the ligand binding site or binding pocket. For example, the compound may be a receptor for KPP which is capable of propagating a signal, or a decoy receptor for KPP which is not capable of propagating a signal (Ashkenazi, A. and V.M. Divit (1999) Curr. Opin. Cell Biol. 11:255-260; Mantovani, A. et al. (2001) Trends Immunol. 22:328-336). The compound can be rationally designed using known techniques. Examples of such techniques include those used to construct the compound etanercept (ENBREL; Immunex Corp., Seattle WA), which is efficacious for treating rheumatoid arthritis in humans. Etanercept is an engineered p75 tumor necrosis factor (TNF) receptor dimer linked to the Fc portion of human IgG₁ (Taylor, P.C. et al. (2001) Curr. Opin. Immunol. 13:611-616).

In one embodiment, screening for compounds which specifically bind to, stimulate, or inhibit KPP involves producing appropriate cells which express KPP, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing KPP or cell membrane fractions which contain KPP are then contacted with a test compound and binding, stimulation, or inhibition of activity of either KPP or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with KPP, either in solution or affixed to a solid support, and detecting the binding of KPP to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

An assay can be used to assess the ability of a compound to bind to its natural ligand and/or to inhibit the binding of its natural ligand to its natural receptors. Examples of such assays include radio-labeling assays such as those described in U.S. Patent No. 5,914,236 and U.S. Patent No. 6,372,724. In a related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a receptor) to improve or alter its ability to bind to its natural ligands. (See, e.g., Matthews, D.J. and J.A. Wells. (1994) Chem. Biol. 1:25-30.) In another related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a ligand) to improve or alter its ability to bind to its natural receptors. (See, e.g., Cunningham, B.C. and J.A. Wells (1991) Proc. Natl. Acad. Sci. USA 88:3407-3411; Lowman, H.B.

et al. (1991) J. Biol. Chem. 266:10982-10988.)

KPP of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of KPP. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for KPP activity, wherein KPP is combined with at least one test compound, and the activity of KPP in the presence of a test compound is compared with the activity of KPP in the absence of the test compound. A change in the activity of KPP in the presence of the test compound is indicative of a compound that modulates the activity of KPP. Alternatively, a test compound is combined with an *in vitro* or cell-free system comprising KPP under conditions suitable for KPP activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of KPP may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding KPP or their mammalian homologs may be “knocked out” in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding KPP may also be manipulated *in vitro* in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding KPP can also be used to create “knockin” humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding KPP is injected into animal ES cells, and the injected sequence

integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress KPP, e.g., by secreting KPP in its milk, may also
 5 serve as a convenient source of that protein (Janne, J. et al. (1998) *Biotechnol. Annu. Rev.* 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of KPP and kinases and phosphatases. In addition, examples of tissues expressing KPP can be found in Table 6. Therefore, KPP appears to play a role in cardiovascular diseases,
 10 immune system disorders, neurological disorders, disorders affecting growth and development, lipid disorders, cell proliferative disorders, and cancers. In the treatment of disorders associated with increased KPP expression or activity, it is desirable to decrease the expression or activity of KPP. In the treatment of disorders associated with decreased KPP expression or activity, it is desirable to increase the expression or activity of KPP.

15 Therefore, in one embodiment, KPP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of KPP. Examples of such disorders include, but are not limited to, a cardiovascular disease such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and
 20 complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis
 25 of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease,
 30 emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary
 35 hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar

proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung transplantation; an immune system disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a disorder affecting growth and development such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis,

primary thrombocythemia, renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary

5 keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a lipid disorder such as fatty liver, cholestasis, primary biliary cirrhosis, carnitine deficiency, carnitine palmitoyltransferase deficiency, myoadenylate deaminase deficiency,

10 hypertriglyceridemia, lipid storage disorders such as Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM₂ gangliosidosis, and ceroid lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, diabetes mellitus, lipodystrophy, lipomatosis, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipid adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia,

15 hypercholesterolemia with hypertriglyceridemia, primary hypoalphalipoproteinemia, hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-Sachs disease, Sandhoff's disease, hyperlipidemia, hyperlipemia, lipid myopathies, and obesity; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective

20 tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen,

25 testis, thymus, thyroid, uterus, leukemias such as multiple myeloma, and lymphomas such as Hodgkin's disease.

In another embodiment, a vector capable of expressing KPP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of KPP including, but not limited to, those described above.

30 In a further embodiment, a composition comprising a substantially purified KPP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of KPP including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of KPP may be

35 administered to a subject to treat or prevent a disorder associated with decreased expression or

activity of KPP including, but not limited to, those listed above.

In a further embodiment, an antagonist of KPP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of KPP. Examples of such disorders include, but are not limited to, those cardiovascular diseases, immune system disorders, 5 neurological disorders, disorders affecting growth and development, lipid disorders, cell proliferative disorders, and cancers described above. In one aspect, an antibody which specifically binds KPP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express KPP.

In an additional embodiment, a vector expressing the complement of the polynucleotide 10 encoding KPP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of KPP including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made 15 by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of KPP may be produced using methods which are generally known in the art. 20 In particular, purified KPP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind KPP. Antibodies to KPP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer 25 formation) are generally preferred for therapeutic use. Single chain antibodies (e.g., from camels or llamas) may be potent enzyme inhibitors and may have advantages in the design of peptide mimetics, and in the development of immuno-adsorbents and biosensors (Muyldermans, S. (2001) J. Biotechnol. 74:277-302).

For the production of antibodies, various hosts including goats, rabbits, rats, mice, camels, 30 dromedaries, llamas, humans, and others may be immunized by injection with KPP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol.

35 Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and *Corynebacterium parvum* are

especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to KPP have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are
5 identical to a portion of the amino acid sequence of the natural protein. Short stretches of KPP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to KPP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not
10 limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the
15 splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce KPP-specific single
20 chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as
25 disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for KPP may also be generated. For example, such fragments include, but are not limited to, F(ab')₂ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of
30 the F(ab')₂ fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either
35 polyclonal or monoclonal antibodies with established specificities are well known in the art. Such

immunoassays typically involve the measurement of complex formation between KPP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering KPP epitopes is generally used, but a competitive binding assay may also be employed (Pound, *supra*).

5 Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for KPP. Affinity is expressed as an association constant, K_a , which is defined as the molar concentration of KPP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their
10 affinities for multiple KPP epitopes, represents the average affinity, or avidity, of the antibodies for KPP. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular KPP epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the KPP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations
15 with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of KPP, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

20 The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of KPP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and
25 guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, *supra*, and Coligan et al. *supra*.)

 In another embodiment of the invention, the polynucleotides encoding KPP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA,
30 RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding KPP. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding KPP. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

 In therapeutic use, any gene delivery system suitable for introduction of the antisense
35 sequences into appropriate target cells can be used. Antisense sequences can be delivered

intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) *J. Allergy Clin. Immunol.* 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral
 5 vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) *Blood* 76:271; Ausubel, *supra*; Uckert, W. and W. Walther (1994) *Pharmacol. Ther.* 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) *Br. Med. Bull.* 51(1):217-225; Boado, R.J. et al. (1998) *J. Pharm. Sci.* 87(11):1308-1315; and Morris, M.C. et al. (1997) *Nucleic Acids Res.*
 10 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding KPP may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) *Science* 288:669-672), severe combined
 15 immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) *Science* 270:475-480; Bordignon, C. et al. (1995) *Science* 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) *Cell* 75:207-216; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:643-666; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:667-703), thalassemias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal,
 20 R.G. (1995) *Science* 270:404-410; Verma, I.M. and N. Somia (1997) *Nature* 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) *Nature* 335:395-396; Poeschla, E. et al. (1996) *Proc. Natl. Acad. Sci. USA* 93:11395-11399),
 25 hepatitis B or C virus (HBV, HCV); fungal parasites, such as *Candida albicans* and *Paracoccidioides brasiliensis*; and protozoan parasites such as *Plasmodium falciparum* and *Trypanosoma cruzi*). In the case where a genetic deficiency in KPP expression or regulation causes disease, the expression of KPP from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in KPP are treated by constructing mammalian expression vectors encoding KPP and introducing these vectors by mechanical means into KPP-deficient cells. Mechanical transfer technologies for use with cells *in vivo* or *ex vitro* include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v)
 35 the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) *Annu. Rev. Biochem.* 62:191-

217; Ivics, Z. (1997) *Cell* 91:501-510; Boulay, J-L. and H. Récipon (1998) *Curr. Opin. Biotechnol.* 9:445-450).

Expression vectors that may be effective for the expression of KPP include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors
 5 (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). KPP may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) *Proc. Natl.*
 10 *Acad. Sci. USA* 89:5547-5551; Gossen, M. et al. (1995) *Science* 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) *Curr. Opin. Biotechnol.* 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, *supra*), or (iii) a tissue-specific promoter or the native promoter of
 15 the endogenous gene encoding KPP from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method
 20 (Graham, F.L. and A.J. Eb (1973) *Virology* 52:456-467), or by electroporation (Neumann, E. et al. (1982) *EMBO J.* 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to KPP expression are treated by constructing a retrovirus vector consisting of (i) the
 25 polynucleotide encoding KPP under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) *Proc.*
 30 *Natl. Acad. Sci. USA* 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) *J. Virol.* 61:1647-1650; Bender, M.A. et al. (1987) *J. Virol.* 61:1639-1646; Adam, M.A. and A.D. Miller (1988) *J. Virol.* 62:3802-3806; Dull, T. et al. (1998) *J. Virol.* 72:8463-8471; Zufferey, R.
 35 et al. (1998) *J. Virol.* 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining

retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant”) discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4⁺ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding KPP to cells which have one or more genetic abnormalities with respect to the expression of KPP. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano (“Adenovirus vectors for gene therapy”), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding KPP to target cells which have one or more genetic abnormalities with respect to the expression of KPP. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing KPP to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca (“Herpes simplex virus strains for gene transfer”), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of

herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding KPP to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for KPP into the alphavirus genome in place of the capsid-coding region results in the production of a large number of KPP-coding RNAs and the synthesis of high levels of KPP in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of KPP into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding KPP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by

scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA sequences encoding KPP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding KPP. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased KPP expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding KPP may be therapeutically useful, and in the treatment of disorders associated with decreased KPP expression or activity, a compound which specifically promotes expression of the polynucleotide encoding KPP may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in

altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound
5 based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding KPP is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an *in vitro* cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding KPP are assayed by
10 any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding KPP. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to
15 a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a *Schizosaccharomyces pombe* gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a
20 combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

25 Many methods for introducing vectors into cells or tissues are available and equally suitable for use *in vivo*, *in vitro*, and *ex vivo*. For *ex vivo* therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections; or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat.
30 Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition
35 which generally comprises an active ingredient formulated with a pharmaceutically acceptable

excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of KPP, antibodies to KPP, and mimetics, agonists, antagonists, or inhibitors of KPP.

5 The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

 Compositions for pulmonary administration may be prepared in liquid or dry powder form. 10 These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, 15 J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

 Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

20 Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising KPP or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, KPP or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to 25 transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

 For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration 30 range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

 A therapeutically effective dose refers to that amount of active ingredient, for example KPP or fragments thereof, antibodies of KPP, and agonists, antagonists or inhibitors of KPP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by 35 standard pharmaceutical procedures in cell cultures or with experimental animals, such as by

calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are
5 used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the
10 subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week,
15 or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg , up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their
20 inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind KPP may be used for the diagnosis of disorders characterized by expression of KPP, or in assays to monitor patients being
25 treated with KPP or agonists, antagonists, or inhibitors of KPP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for KPP include methods which utilize the antibody and a label to detect KPP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter
30 molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring KPP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of KPP expression. Normal or standard values for KPP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to KPP under
35 conditions suitable for complex formation. The amount of standard complex formation may be

quantitated by various methods, such as photometric means. Quantities of KPP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding KPP may be used for
5 diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of KPP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of KPP, and to monitor regulation of KPP levels during therapeutic intervention.

10 In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding KPP or closely related molecules may be used to identify nucleic acid sequences which encode KPP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the
15 probe identifies only naturally occurring sequences encoding KPP, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the KPP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:15-28 or from
20 genomic sequences including promoters, enhancers, and introns of the KPP gene.

Means for producing specific hybridization probes for DNAs encoding KPP include the cloning of polynucleotide sequences encoding KPP or KPP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerases and the
25 appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ³²P or ³⁵S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding KPP may be used for the diagnosis of disorders associated with expression of KPP. Examples of such disorders include, but are not limited to, a
30 cardiovascular disease such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular
35 heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular

calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung transplantation; an immune system disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal

familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia,

5 Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a disorder affecting growth and development such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism,

10 Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida,

15 anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a lipid disorder such as fatty liver, cholestasis, primary biliary cirrhosis, carnitine deficiency, carnitine palmitoyltransferase deficiency, myoadenylate deaminase deficiency, hypertriglyceridemia, lipid storage disorders such as Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM₂ gangliosidosis, and ceroid lipofuscinosis,

20 abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, diabetes mellitus, lipodystrophy, lipomatosis, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipoid adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with hypertriglyceridemia, primary hypoalphalipoproteinemia, hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous

25 xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-Sachs disease, Sandhoff's disease, hyperlipidemia, hyperlipemia, lipid myopathies, and obesity; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma,

30 melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland,

bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, uterus, leukemias such as multiple myeloma, and lymphomas such as Hodgkin's disease. The polynucleotide sequences encoding KPP may be used in Southern or
5 northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered KPP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding KPP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide
10 sequences encoding KPP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding KPP in the
15 sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of KPP, a normal or standard profile for expression is established. This may be accomplished by combining
20 body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding KPP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from
25 samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from
30 successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance
35 of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals

to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding KPP may involve the use of PCR. These oligomers may be chemically synthesized, generated
5 enzymatically, or produced *in vitro*. Oligomers will preferably contain a fragment of a polynucleotide encoding KPP, or a fragment of a polynucleotide complementary to the polynucleotide encoding KPP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

10 In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding KPP may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP,
15 oligonucleotide primers derived from the polynucleotide sequences encoding KPP are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSSCP, the
20 oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed *in silico* SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation
25 of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

SNPs may be used to study the genetic basis of human disease. For example, at least 16 common SNPs have been associated with non-insulin-dependent diabetes mellitus. SNPs are also
30 useful for examining differences in disease outcomes in monogenic disorders, such as cystic fibrosis, sickle cell anemia, or chronic granulomatous disease. For example, variants in the mannose-binding lectin, MBL2, have been shown to be correlated with deleterious pulmonary outcomes in cystic fibrosis. SNPs also have utility in pharmacogenomics, the identification of genetic variants that influence a patient's response to a drug, such as life-threatening toxicity. For example, a variation in
35 N-acetyl transferase is associated with a high incidence of peripheral neuropathy in response to the

anti-tuberculosis drug isoniazid, while a variation in the core promoter of the ALOX5 gene results in diminished clinical response to treatment with an anti-asthma drug that targets the 5-lipoxygenase pathway. Analysis of the distribution of SNPs in different populations is useful for investigating genetic drift, mutation, recombination, and selection, as well as for tracing the origins of populations and their migrations. (Taylor, J.G. et al. (2001) Trends Mol. Med. 7:507-512; Kwok, P.-Y. and Z. Gu (1999) Mol. Med. Today 5:538-543; Nowotny, P. et al. (2001) Curr. Opin. Neurobiol. 11:637-641.)

Methods which may also be used to quantify the expression of KPP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, KPP, fragments of KPP, or antibodies specific for KPP may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of

transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

5 Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression *in vivo*, as in the case of a tissue or biopsy sample, or *in vitro*, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with *in vitro* model systems and preclinical evaluation of
10 pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471). If a test compound has a signature similar to that of a
15 compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression
20 data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29,
25 2000, available at <http://www.niehs.nih.gov/oc/news/toxchip.htm>.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of
30 the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present
35 invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global

pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for KPP to quantify the levels of KPP expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) *Anal. Biochem.* 270:103-111; Mendoze, L.G. et al. (1999) *Biotechniques* 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) *Electrophoresis* 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London.

In another embodiment of the invention, nucleic acid sequences encoding KPP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP).

(See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent *in situ* hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, *supra*, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding KPP on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, KPP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between KPP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with KPP, or fragments thereof, and washed. Bound KPP is then detected by methods well known in the art. Purified KPP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding KPP specifically compete with a test compound for binding KPP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with KPP.

In additional embodiments, the nucleotide sequences which encode KPP may be used in any

molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding
5 description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications and publications, mentioned above and below, including U.S. Ser. No. 60/293,665, U.S. Ser. No. 60/298,712, U.S. Ser. No. 60/303,418, U.S. Ser.
10 No. 60/306,967, U.S. Ser. No. 60/308,183, U.S. Ser. No. 60/343,007, U.S. Ser. No. 60/357,675, and U.S. Ser. No. 60/376,988, are expressly incorporated by reference herein.

EXAMPLES

I. Construction of cDNA Libraries

15 Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA). Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Invitrogen), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with
20 chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN,
25 Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP
30 vector system (Stratagene) or SUPERSCRIPT plasmid system (Invitrogen), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, *supra*, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using
35 SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography

(Amersham Biosciences) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Invitrogen), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte Genomics, Palo Alto CA), pRARE (Incyte Genomics), or pINCY (Incyte Genomics), or derivatives thereof. Recombinant plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 α , DH10B, or ElectroMAX DH10B from Invitrogen.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by *in vivo* excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Biosciences or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Amersham Biosciences); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel,

1997, *supra*, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from *Homo sapiens*, *Rattus norvegicus*, *Mus musculus*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and *Candida albicans* (Incyte Genomics, Palo Alto CA); hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM (Haft, D.H. et al. (2001) Nucleic Acids Res. 29:41-43); and HMM-based protein domain databases such as SMART (Schultz et al. (1998) Proc. Natl. Acad. Sci. USA 95:5857-5864; Letunic, I. et al. (2002) Nucleic Acids Res. 30:242-244). (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM; and HMM-based protein domain databases such as SMART. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and

threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:15-28. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative kinases and phosphatases were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode kinases and phosphatases, the encoded polypeptides were analyzed by querying against PFAM models for kinases and phosphatases. Potential kinases and phosphatases were also identified by homology to Incyte cDNA sequences that had been annotated as kinases and phosphatases. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbprc public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

20 "Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of KPP Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:15-28 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other

implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:15-28 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for
 5 Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-
 10 arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and
 15 other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (<http://www.ncbi.nlm.nih.gov/genemap/>), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a
 20 gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, *supra*, ch. 7; Ausubel (1995) *supra*, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is
 25 much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum \{length(Seq. 1), length(Seq. 2)\}}}$$

30

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the
 35 product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is

calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding KPP are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding KPP. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of KPP Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+} , $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Biosciences), ELONGASE enzyme (Invitrogen), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 μ l PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 μ l of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μ l to 10 μ l aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviII cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Biosciences). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Biosciences), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Biosciences) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers

and the DYENAMIC DIRECT kit (Amersham Biosciences) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides
5 designed for such extension, and an appropriate genomic library.

IX. Identification of Single Nucleotide Polymorphisms in KPP Encoding Polynucleotides

Common DNA sequence variants known as single nucleotide polymorphisms (SNPs) were identified in SEQ ID NO:15-28 using the LIFESEQ database (Incyte Genomics). Sequences from the same gene were clustered together and assembled as described in Example III, allowing the
10 identification of all sequence variants in the gene. An algorithm consisting of a series of filters was used to distinguish SNPs from other sequence variants. Preliminary filters removed the majority of basecall errors by requiring a minimum Phred quality score of 15, and removed sequence alignment errors and errors resulting from improper trimming of vector sequences, chimeras, and splice variants. An automated procedure of advanced chromosome analysis analysed the original
15 chromatogram files in the vicinity of the putative SNP. Clone error filters used statistically generated algorithms to identify errors introduced during laboratory processing, such as those caused by reverse transcriptase, polymerase, or somatic mutation. Clustering error filters used statistically generated algorithms to identify errors resulting from clustering of close homologs or pseudogenes, or due to contamination by non-human sequences. A final set of filters removed duplicates and SNPs found in
20 immunoglobulins or T-cell receptors.

Certain SNPs were selected for further characterization by mass spectrometry using the high throughput MASSARRAY system (Sequenom, Inc.) to analyze allele frequencies at the SNP sites in four different human populations. The Caucasian population comprised 92 individuals (46 male, 46 female), including 83 from Utah, four French, three Venezuelan, and two Amish individuals. The
25 African population comprised 194 individuals (97 male, 97 female), all African Americans. The Hispanic population comprised 324 individuals (162 male, 162 female), all Mexican Hispanic. The Asian population comprised 126 individuals (64 male, 62 female) with a reported parental breakdown of 43% Chinese, 31% Japanese, 13% Korean, 5% Vietnamese, and 8% other Asian. Allele frequencies were first analyzed in the Caucasian population; in some cases those SNPs which showed
30 no allelic variance in this population were not further tested in the other three populations.

X. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:15-28 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide
35 fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06

software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μ Ci of [γ - 32 P] adenosine triphosphate (Amersham Biosciences), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Biosciences). An aliquot containing 10⁷ counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

XI. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, *supra.*), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), *supra.*). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) *Science* 270:467-470; Shalon, D. et al. (1996) *Genome Res.* 6:639-645; Marshall, A. and J. Hodgson (1998) *Nat. Biotechnol.* 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is

described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)⁺ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)⁺ RNA sample is
5 reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ μ l oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/ μ l RNase inhibitor, 500 μ M dATP, 500 μ M dGTP, 500 μ M dTTP, 40 μ M dCTP, 40 μ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Biosciences). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)⁺ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)⁺ RNAs are synthesized by *in vitro* transcription
10 from non-coding yeast genomic DNA. After incubation at 37° C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85° C to stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated
15 using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μ l 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element
20 is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μ g. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Biosciences).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope
25 slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

30 Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1 μ l of the array element DNA, at an average concentration of 100 ng/ μ l, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene).
35 Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water.

Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60° C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

5 Hybridization reactions contain 9 μ l of sample mixture consisting of 0.2 μ g each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65° C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm² coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the
10 addition of 140 μ l of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60° C. The arrays are washed for 10 min at 45° C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45° C in a second wash buffer (0.1X SSC), and dried.

Detection

15 Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-
20 scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

 In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate
25 filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

 The sensitivity of the scans is typically calibrated using the signal intensity generated by a
30 cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are
35 differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the

two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a
5 linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each
10 spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

XII. Complementary Polynucleotides

Sequences complementary to the KPP-encoding sequences, or any parts thereof, are used to
15 detect, decrease, or inhibit expression of naturally occurring KPP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of KPP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent
20 promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the KPP-encoding transcript.

XIII. Expression of KPP

Expression and purification of KPP is achieved using bacterial or virus-based expression systems. For expression of KPP in bacteria, cDNA is subcloned into an appropriate vector containing
25 an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac (tac)* hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express KPP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG).
30 Expression of KPP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding KPP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels
35 of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect

cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, KPP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Biosciences). Following purification, the GST moiety can be proteolytically cleaved from KPP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, *supra*, ch. 10 and 16). Purified KPP obtained by these methods can be used directly in the assays shown in Examples XVII, XVIII, XIX, XX, and XXI, where applicable.

XIV. Functional Assays

KPP function is assessed by expressing the sequences encoding KPP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT plasmid (Invitrogen, Carlsbad CA) and PCR3.1 plasmid (Invitrogen), both of which contain the cytomegalovirus promoter. 5-10 μ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding

of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of KPP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding KPP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding KPP and other genes of interest can be analyzed by northern analysis or microarray techniques.

XV. Production of KPP Specific Antibodies

KPP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize animals (e.g., rabbits, mice, etc.) and to produce antibodies using standard protocols.

Alternatively, the KPP amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, *supra*, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, *supra*.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for anti-peptide and anti-KPP activity by, for example, binding the peptide or KPP to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XVI. Purification of Naturally Occurring KPP Using Specific Antibodies

Naturally occurring or recombinant KPP is substantially purified by immunoaffinity chromatography using antibodies specific for KPP. An immunoaffinity column is constructed by covalently coupling anti-KPP antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Biosciences). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing KPP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of KPP (e.g., high ionic strength buffers in the

presence of detergent). The column is eluted under conditions that disrupt antibody/KPP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and KPP is collected.

XVII. Identification of Molecules Which Interact with KPP

5 KPP, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton, A.E. and W.M. Hunter (1973) *Biochem. J.* 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled KPP, washed, and any wells with labeled KPP complex are assayed. Data obtained using different concentrations of KPP are used to calculate values for the number, affinity, and association of KPP with the candidate
10 molecules.

Alternatively, molecules interacting with KPP are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) *Nature* 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

KPP may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT)
15 which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVIII. Demonstration of KPP Activity

Generally, protein kinase activity is measured by quantifying the phosphorylation of a protein
20 substrate by KPP in the presence of $[\gamma\text{-}^{32}\text{P}]\text{ATP}$. KPP is incubated with the protein substrate, ^{32}P -ATP, and an appropriate kinase buffer. The ^{32}P incorporated into the substrate is separated from free ^{32}P -ATP by electrophoresis and the incorporated ^{32}P is counted using a radioisotope counter. The amount of incorporated ^{32}P is proportional to the activity of KPP. A determination of the specific amino acid residue phosphorylated is made by phosphoamino acid analysis of the hydrolyzed protein.

25 In one alternative, protein kinase activity is measured by quantifying the transfer of gamma phosphate from adenosine triphosphate (ATP) to a serine, threonine or tyrosine residue in a protein substrate. The reaction occurs between a protein kinase sample with a biotinylated peptide substrate and gamma ^{32}P -ATP. Following the reaction, free avidin in solution is added for binding to the biotinylated ^{32}P -peptide product. The binding sample then undergoes a centrifugal ultrafiltration
30 process with a membrane which will retain the product-avidin complex and allow passage of free gamma ^{32}P -ATP. The reservoir of the centrifuged unit containing the ^{32}P -peptide product as retentate is then counted in a scintillation counter. This procedure allows the assay of any type of protein kinase sample, depending on the peptide substrate and kinase reaction buffer selected. This assay is provided in kit form (ASUA, Affinity Ultrafiltration Separation Assay, Transbio Corporation,
35 Baltimore MD, U.S. Patent No. 5,869,275). Suggested substrates and their respective enzymes

include but are not limited to: Histone H1 (Sigma) and p34^{cdc2}kinase, Annexin I, Angiotensin (Sigma) and EGF receptor kinase, Annexin II and *src* kinase, ERK1 & ERK2 substrates and MEK, and myelin basic protein and ERK (Pearson, J.D. et al. (1991) Methods Enzymol. 200:62-81).

In another alternative, protein kinase activity of KPP is demonstrated in an assay containing
5 KPP, 50 μ l of kinase buffer, 1 μ g substrate, such as myelin basic protein (MBP) or synthetic peptide substrates, 1 mM DTT, 10 μ g ATP, and 0.5 μ Ci [γ -³²P]ATP. The reaction is incubated at 30°C for 30 minutes and stopped by pipetting onto P81 paper. The unincorporated [γ -³²P]ATP is removed by washing and the incorporated radioactivity is measured using a scintillation counter. Alternatively, the reaction is stopped by heating to 100°C in the presence of SDS loading buffer and resolved on a
10 12% SDS polyacrylamide gel followed by autoradiography. The amount of incorporated ³²P is proportional to the activity of KPP.

In yet another alternative, adenylate kinase or guanylate kinase activity of KPP may be measured by the incorporation of ³²P from [γ -³²P]ATP into ADP or GDP using a gamma radioisotope counter. KPP, in a kinase buffer, is incubated together with the appropriate nucleotide
15 mono-phosphate substrate (AMP or GMP) and ³²P-labeled ATP as the phosphate donor. The reaction is incubated at 37°C and terminated by addition of trichloroacetic acid. The acid extract is neutralized and subjected to gel electrophoresis to separate the mono-, di-, and triphosphonucleotide fractions. The diphosphonucleotide fraction is excised and counted. The radioactivity recovered is proportional to the activity of KPP.

20 In yet another alternative, other assays for KPP include scintillation proximity assays (SPA), scintillation plate technology and filter binding assays. Useful substrates include recombinant proteins tagged with glutathione transferase, or synthetic peptide substrates tagged with biotin. Inhibitors of KPP activity, such as small organic molecules, proteins or peptides, may be identified by such assays.

25 In another alternative, phosphatase activity of KPP is measured by the hydrolysis of para-nitrophenyl phosphate (PNPP). KPP is incubated together with PNPP in HEPES buffer pH 7.5, in the presence of 0.1% β -mercaptoethanol at 37°C for 60 min. The reaction is stopped by the addition of 6 ml of 10 N NaOH (Diamond, R.H. et al. (1994) Mol. Cell. Biol. 14:3752-62). Alternatively, acid phosphatase activity of KPP is demonstrated by incubating KPP-containing extract with 100 μ l of 10
30 mM PNPP in 0.1 M sodium citrate, pH 4.5, and 50 μ l of 40 mM NaCl at 37°C for 20 min. The reaction is stopped by the addition of 0.5 ml of 0.4 M glycine/NaOH, pH 10.4 (Saftig, P. et al. (1997) J. Biol. Chem. 272:18628-18635). The increase in light absorbance at 410 nm resulting from the hydrolysis of PNPP is measured using a spectrophotometer. The increase in light absorbance is proportional to the activity of KPP in the assay.

35 In the alternative, KPP activity is determined by measuring the amount of phosphate removed

from a phosphorylated protein substrate. Reactions are performed with 2 or 4 nM KPP in a final volume of 30 μ l containing 60 mM Tris, pH 7.6, 1 mM EDTA, 1 mM EGTA, 0.1% β -mercaptoethanol and 10 μ M substrate, 32 P-labeled on serine/threonine or tyrosine, as appropriate. Reactions are initiated with substrate and incubated at 30° C for 10-15 min. Reactions are quenched with 450 μ l of
5 4% (w/v) activated charcoal in 0.6 M HCl, 90 mM $\text{Na}_4\text{P}_2\text{O}_7$, and 2 mM NaH_2PO_4 , then centrifuged at 12,000 $\times g$ for 5 min. Acid-soluble ^{32}P i is quantified by liquid scintillation counting (Sinclair, C. et al. (1999) J. Biol. Chem. 274:23666-23672).

XIX. Kinase Binding Assay

Binding of KPP to a FLAG-CD44 cyt fusion protein can be determined by incubating KPP
10 with anti-KPP-conjugated immunoaffinity beads followed by incubating portions of the beads (having 10-20 ng of protein) with 0.5 ml of a binding buffer (20 mM Tris-HCL (pH 7.4), 150 mM NaCl, 0.1% bovine serum albumin, and 0.05% Triton X-100) in the presence of ^{125}I -labeled FLAG-CD44cyt fusion protein (5,000 cpm/ng protein) at 4 °C for 5 hours. Following binding, beads were washed thoroughly in the binding buffer and the bead-bound radioactivity measured in a scintillation counter
15 (Bourguignon, L.Y.W. et al. (2001) J. Biol. Chem. 276:7327-7336). The amount of incorporated ^{32}P is proportional to the amount of bound KPP.

XX. Identification of KPP Inhibitors

Compounds to be tested are arrayed in the wells of a 384-well plate in varying concentrations along with an appropriate buffer and substrate, as described in the assays in Example XVII. KPP
20 activity is measured for each well and the ability of each compound to inhibit KPP activity can be determined, as well as the dose-response kinetics. This assay could also be used to identify molecules which enhance KPP activity.

XXI. Identification of KPP Substrates

A KPP "substrate-trapping" assay takes advantage of the increased substrate affinity that may
25 be conferred by certain mutations in the PTP signature sequence of protein tyrosine phosphatases. KPP bearing these mutations form a stable complex with their substrate; this complex may be isolated biochemically. Site-directed mutagenesis of invariant residues in the PTP signature sequence in a clone encoding the catalytic domain of KPP is performed using a method standard in the art or a commercial kit, such as the MUTA-GENE kit from BIO-RAD. For expression of KPP mutants in
30 *Escherichia coli*, DNA fragments containing the mutation are exchanged with the corresponding wild-type sequence in an expression vector bearing the sequence encoding KPP or a glutathione S-transferase (GST)-KPP fusion protein. KPP mutants are expressed in *E. coli* and purified by chromatography.

The expression vector is transfected into COS1 or 293 cells via calcium phosphate-mediated
35 transfection with 20 μ g of CsCl-purified DNA per 10-cm dish of cells or 8 μ g per 6-cm dish. Forty-

eight hours after transfection, cells are stimulated with 100 ng/ml epidermal growth factor to increase tyrosine phosphorylation in cells, as the tyrosine kinase EGFR is abundant in COS cells. Cells are lysed in 50 mM Tris·HCl, pH 7.5/5 mM EDTA/150 mM NaCl/1% Triton X-100/5 mM iodoacetic acid/10 mM sodium phosphate/10 mM NaF/5 μ g/ml leupeptin/5 μ g/ml aprotinin/1 mM benzamidine
5 (1 ml per 10-cm dish, 0.5 ml per 6-cm dish). KPP is immunoprecipitated from lysates with an appropriate antibody. GST-KPP fusion proteins are precipitated with glutathione-Sepharose, 4 μ g of mAb or 10 μ l of beads respectively per mg of cell lysate. Complexes can be visualized by PAGE or further purified to identify substrate molecules (Flint, A.J. et al. (1997) Proc. Natl. Acad. Sci. USA 94:1680-1685).

10

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific
15 embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Incyte Full Length Clones
2537210	1	2537210CD1	15	2537210CB1	
112535	2	112535CD1	16	112535CB1	90099938CA2
72063274	3	72063274CD1	17	72063274CB1	90065795CA2
5013673	4	5013673CD1	18	5013673CB1	90099936CA2
5977982	5	5977982CD1	19	5977982CB1	5977982CA2
6880271	6	6880271CD1	20	6880271CB1	
2378756	7	2378756CD1	21	2378756CB1	90097266CA2
1861527	8	1861527CD1	22	1861527CB1	
2921356	9	2921356CD1	23	2921356CB1	
7386170	10	7386170CD1	24	7386170CB1	
7481206	11	7481206CD1	25	7481206CB1	
7503117	12	7503117CD1	26	7503117CB1	
7506911	13	7506911CD1	27	7506911CB1	90116059CA2
7510809	14	7510809CD1	28	7510809CB1	

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability Score	GenBank Homolog
1	2537210CD1	g1545959	0.0	[Mus musculus] paladin
2	112535CD1	g6580857	0.0	serine/threonine kinase UNC51.2 [Mus musculus] (Tomada, T. et al. (1999) Neuron 24:833-846)
3	72063274CD1	g10177920	2.5E-33	[Arabidopsis thaliana] contains similarity to adenylate kinase gene_id:MCA23.18
4	5013673CD1	g3702174	5.7E-90	[Mus musculus] Fish protein (Lock, P. et al. (1998) EMBO J. 17:4346-4357)
5	5977982CD1	g2052058	1.1E-188	[Homo sapiens] SIRP-beta1 (Kharitonov, A. et al. (1997) Nature 386:181-186)
6	6880271CD1	g13162658	2.2E-183	[Rattus norvegicus] inositol polyphosphate multikinase (Saiardi, A. et al. (2001) Proc. Natl. Acad. Sci. USA 98:2306-2311)
8	1861527CD1	g15131540	0.0	serine/threonine protein kinase [Homo sapiens]
9	2921356CD1	g6690020	1.4E-194	[Mus musculus] pantothenate kinase 1 beta (Rock, C.O. et al. (2000) J. Biol. Chem. 275:1377-1383)
10	7386170CD1	g35501	3.3E-296	[Homo sapiens] protein kinase C zeta (Kochs, G. et al. (1993) Eur. J. Biochem. 216:597-606)
11	7481206CD1	g1502311	0.0	[Homo sapiens] phosphorylase kinase (Wullrich-Schmoll, A. et al. (1996) Eur. J. Biochem. 38:374-380)
12	7503117CD1	g4760561	0.0	[Mus musculus] UNC-51-like kinase (ULK) 2 (Yan, J. et al. (1999) Oncogene 18:5850-5859)
		346446 K1AA0623	0.0	[Homo sapiens] [Protein kinase; transferase] Protein with very strong similarity to murine Ulk2, which is a member of the Unc-51 like family of serine/threonine protein kinases; expressed in brain
		438255 ULK2	0.0	[Mus musculus] [Protein kinase; transferase] Member of the UNC-51-like family of serine/threonine protein kinases (Yan, J. et al. (1999) supra)
13	7506911CD1	g6690020	1.7E-196	[Mus musculus] pantothenate kinase 1 beta (Rock, C.O. et al. (2000) supra)
		690996 FLJ12899	1.0E-158	[Homo sapiens] Protein with low similarity to S. cerevisiae Ydr531p, which is a putative pantothenate kinase involved in coenzyme A biosynthesis

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability Score	GenBank Homolog
14	7510809CD1	370685 SPBC428.19c	1.5E-40	[Schizosaccharomyces pombe] WD-repeat protein
		634688 orf6.2465	4.8E-28	[Candida albicans] Protein containing six WD domains (WD-40 repeat), which likely mediate protein-protein interactions, has a region of low similarity to a region of murine Tle4, which is a Groucho-related protein (Huang, S. et al. (1987) Biochemistry 26:8242-8246)
		742592 DKFZP434C245	5.3E-19	[Homo sapiens] Protein containing WD domains (WD-40 repeat), which may mediate protein-protein interactions

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	2537210CD1	856	S276 S338 S483 S521 S546 S561 S695 S827 S850 T16 T106 T181 T263 T497 T597 T605 T702 T732 T751 Y180		Transmembrane domain: V316-R333, G666-I682 N-terminus is non-cytosolic PALADIN GENE PD148197: M1-L856	TMAP BLAST- PRODOM
2	112535CD1	1036	S10 S147 S148 S218 S301 S309 S342 S346 S378 S430 S468 S471 S582 S651 S719 S764 S789 S908 S947 S1027 T32 T102 T107 T483 T531 T670 T694 T817 T826 T939	N45 N229 N320 N345 N359 N416 N505	Protein kinase domain: Y9-L271 Protein kinases signatures and profile: T107-M169 Tyrosine kinase catalytic domain signature PR00109: T240-F262, M85-Q98, I121-L139 KINASE ULK1 SERINE/THREONINE UNC51 LIKE UNC51-LIKE TRANSFERASE SERINE/THREONINE PROTEIN ATP-BINDING KIAA0623 PD144878: S393-A810 KINASE ULK1 UNC51 TRANSFERASE SERINE/THREONINE PROTEIN ATP-BINDING SERINE/THREONINE LIKE UNC51-LIKE KIAA0623 PD043441: P811-L1029 PROTEIN KINASE DOMAIN DM00004 P53104 26-315: F111-F262, V15-D106 JC1446 20-261: V15-L252 A53215 585-829: D13-R259 I48719 591-835: D13-R259 Protein kinases ATP-binding region signature: V15-K39 Serine/Threonine protein kinases active-site signature: I127-L139	HMME-PFAM PROFILES BLIMPS- PRINTS BLAST- PRODOM BLAST- PRODOM BLAST-DOMO MOTIFS MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
3	72063274CD1	479	S68 S165 S211 S279 S337 S384 T99 T157 T190 T268 T312 T390 T461 Y17 Y437	N82	Adenylate kinase: L273-P409, V113-F192, I62-K70 Adenylate kinase protein BL00113: N296-R339, C345-L359, L419-S449 Shikimate kinase family signature PR01100: I60-M75, N373-T390 Adenylate kinase signature protein PR00094: V61-A74 KINASE ADENYLATE TRANSFERASE ATP BINDING ATP AMP TRANSPHOSPHORYLASE ISOENZYME PROTEIN 3D STRUCTURE MITOCHONDRION PD000657: L273-Y447, I62-K235 ADENYLATE KINASE DM00290 P24323 1-177: R270-L443, R59-R224 I64062 1-174: R270-P409, R59-R224 P43188 1-128: R270-R392, R59-K178 P27144 1-125: R270-R392 ATP/GTP-binding site motif A (P-loop): G64-T71, G275-S282	HMMER-PFAM BLIMPS-BLOCKS BLIMPS-PRINTS BLIMPS-PRINTS BLAST-PRODROM
4	5013673CD1	760	S42 S128 S168 S188 S278 S348 S361 S377 S381 S424 S467 S498 S528 S569 S574 S618 S620 S718 T43 T53 T82 T224 T231 T269 T308	N34 N250 N279 N306 N731	SH3 domain: E73-N127, E4-Q58, E220-Y274, S702-K759 SH3 domain signature PR00452: S702-G712, S732-L741, W747-K759 KIAA0418 FISH PROTEIN PD147796: P555-K759 KIAA0418 FISH PROTEIN PD044000: Y222-I271	MOTIFS HMMER-PFAM BLIMPS-PRINTS BLAST-PRODROM BLAST-PRODROM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
4			T402 T562 T678 T723		SRC HOMOLOGY 3 (SH3) DOMAIN DM00025 P14598 157-217: V2-E64	BLAST-DOMO
5	5977982CD1	398	S104 S127 S142 S175 S293 T91 T96 T111 T180 T195 T211 T217 T248 T285 T300	N244 N269 N291 N318	signal cleavage: M1-G29 Signal Peptide: M1-G29 Immunoglobulin domain: G47-V121, Q162-V229, E265-Y309, R323-S341 Transmembrane domain: S6-V27, P361-V389 N-terminus is non-cytosolic	SPSCAN HMMER HMMER-PFAM
6	6880271CD1	416	S27 S158 S284 S334 S360 S412 T134 T206 T265 T272 T326 T394 Y90		SIGNAL REGULATORY PROTEIN BETAL PRECURSOR SIRP BETAL SIGNAL IMMUNOGLOBULIN FOLD GLYCOPROTEIN TRANSMEMBRANE PD087237: M1-A46 PD054160: Q337-A398 PROTEIN ARGININE METABOLISM REGULATION III TRANSCRIPTION SIMILARITY SACCHAROMYCES CEREVISIAE PUTATIVE PD011544: Y127-G259	BLAST- PRODOM
7	2378756CD1	359	S42 S59 S110 S139 S243 S279 S314 S326 T87 T145 T159 T188 Y25 Y84		WD domain, G-beta repeat: P114-D150, L240-S276, P72-D108, S199-D233 E156-D193, K30-G66, V282-H317 Trp-Asp (WD-40) repeats signature: Y84-F130, T127-K174 Trp-Asp (WD) repeat proteins proteins BL00678: S139-W149 G-Protein Beta WD-40 repeat signature PR00320: V137-I151 Trp-Asp (WD) repeats signature: V137-I151	HMMER-PFAM PROFILES SCAN BLIMPS- BLOCKS BLIMPS- PRINTS MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
8	1861527CD1	1243	S14 S47 S56	N1112	Protein kinase domain: I174-F432	HMMER-PFAM
			S60 S63 S64		Protein kinases signatures and profile: L277-F332	PROFILES SCAN
			S101 S122		Tyrosine kinase catalytic domain signature PR00109: T254-R267, H292-I310, V356-C378, K401-I423	BLIMPS-PRINTS
			S184 S211		Biotin repressor PF01317: G496-E512	BLIMPS-PFAM
			S239 S316		KINASE WNK-1 KIAA0344 NY-CO-43	BLAST-PRODOR
			S331 S390		ANTIGEN PD041299: S1114-P1207	BLAST-DOMO
			S581 S611		PROTEIN KINASE DOMAIN DM00004	
			S637 S680		S49611 39-259: I180-K401	
			S711 S728		Q05609 553-797: E179-C412	
			S763 S782		P51957 8-251: I180-R420	
9	2921356CD1	373	S868 S1021	N72 N103	P41892 11-249: I180-R420	
			S1041 S1054		Serine/Threonine protein kinases active-site signature: I298-I310	MOTIFS
			S1073 S1121			
			S1127 S1128			
			S1140 S1147			
			S1183 S1190			
			S1197 S1201			
			S1217 T16			
			T109 T167			
			T196 T207			
9	2921356CD1	373	T261 T326	N72 N103	PROTEIN T13D8.31 D9719.34P COSMID	BLAST-PRODOR
			T389 T415		C10G11 CODED FOR BY C ELEGANS (KINASE	
			T422 T579		CDNA PANTOTHENATE BETA FIS SIMILAR	
			T702 T752		MUSCULUS MODERATELY PANK1BET)	
			T790 T1196		PD018089: M16-D288	
			Y1164			
			S3 S47 S100			
			S283 S285			
			S330 T34			
			T167 T209			
9	2921356CD1	373	T226 T230	N72 N103		
			T244			

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
10	7386170CD1	548	S63 S72 S98 S173 S174 S179 S405 S438 S529 S543 T29 T35 T36 T57 T255 T279 T516 Y234		Octicosapeptide repeat: P55-R84 Protein kinase domain: F208-F474 Protein kinase C terminal domain: R475-L541 Tyrosine kinase catalytic domain signature PR00109: I286-Q299, F322-L340, V388-I410, L437-C459 Octicosapeptide repeat p PF00564: H273-L323, E483-Y536, F208-Q262 PROTEIN KINASE C ZETA TYPE NPKCZETA ATPBINDING TRANSFERASE SERINE/THREONINEPROTEIN PHORBOLESTER PD012822: R137-D207 C PROTEIN KINASE PHORBOLESTER BINDING TYPE ATPBINDING TRANSFERASE SERINE/THREONINEPROTEIN ZINC PD150154: R87-G130 PD150295: M1-T57 PROTEIN KINASE DOMAIN DM00004 P09217 254-502: L210-C459 Q05513 246-494: L210-C459 Q02956 254-502: L210-C459 P41743 247-497: L210-C459 Protein kinases ATP-binding region signature: I214-K237 Serine/Threonine protein kinases active-site signature: I328-L340	HMMER-PFAM HMMER-PFAM HMMER-PFAM BLIMPS-PRINTS BLIMPS-PFAM BLAST-PRODOM BLAST-PRODOM BLAST-DOMO MOTIFS MOTIFS
11	7481206CD1	1093	S12 S27 S163 S341 S426 S460 S480 S503 S550 S674 S684	N42 N229 N478 N715 N978	Colicin immunity protein PF01320: V810-Q850, L317-F333	BLIMPS-PFAM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
11			S700 S704 S775 S827 S903 T22 T44 T157 T340 T562 T761 T841 T968 T973 T1065 Y89		PHOSPHORYLASE KINASE ALP PD001841: Q386-K439, E490-I526, V762-V803, Q851-L887, W924-P953, T968-V1009, F1054-A1072, R110-D159, Y181-W221, E222-A242, K243-A294, A295-R332, F333-N381 PHOSPHORYLASE B KINASE REGULATORY CHAIN SUBUNIT GLYCOGEN METABOLISM PHOSPHORYLATION CALMODULIN BINDING (PHOSPHORYLASE B KINASE BETA REGULATORY CHAIN (PHOSPHORYLASE KINASE BETA SUBUNIT)) PD005098: A2-I716	BLIMPS- PRODOM BLAST- PRODOM
					PHOSPHORYLASE B KINASE REGULATORY CHAIN SUBUNIT GLYCOGEN METABOLISM PHOSPHORYLATION CALMODULIN BINDING PD008424: Q918-I1045	BLAST- PRODOM
					CALMODULIN-BINDING DOMAIN DM03490 P12798 36-1091: L37-S1093 S24109 1-1235: K48-R469, G438-S694, Q918-E1079, D721-G921 P34335 61-1256: K48-E707, W924-Y1068 P46020 1-1222: K48-V508, L826-E1079	BLAST-DOMO
12	7503117CD1	1009	S10 S147 S148 S218 S301 S309 S342 S346 S378 S430 S468 S471 S582 S651 S737 S762 S881 S920 S1000 T32	N45 N229 N320 N345 N359 N416 N505	Protein kinase domain: Y9-L271 Protein kinases signatures and profile: T107-M169 Tyrosine kinase catalytic domain signature PR00109: I121-L139, T240-F262, M85-Q98 KINASE ULK1 SERINE/THREONINE UNC51 LIKE UNC51LIKE TRANSFERASE SERINE/THREONINEPROTEIN ATPBINDING KIAA0623 PD144878: S393-P755	HMME-PFAM PROFILESSCAN BLIMPS- PRINTS BLAST- PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
12			T102 T107 T483 T531 T670 T694 T790 T799 T912		KINASE ULK1 UNC51 TRANSFERASE SERINE/THREONINEPROTEIN ATPBINDING SERINE/THREONINE LIKE UNC51LIKE KIAA0623 PD043441: P784-L1002 PROTEIN KINASE DOMAIN MD00004 P53104 26-315: F111-F262, V15-D106 JC1446 20-261: V15-L252 A53215 585-829: D13-R259 I48719 591-835: D13-R259 Protein kinases ATP-binding region signature: V15-K39 Serine/Threonine protein kinases active-site signature: I127-L139 pank_eukar: pantothenate kinase: T66-E344 KINASE TRANSFERASE CDNA PANTOTHENATE FIS FOR T13D8.31 CODED COSMID ELEGANS PD018089: M48-D320, N306-L396	BLAST- PRODOM BLAST-DOMO
13	7506911CD1	405	S13 S79 S132 S315 S317 S362 T11 T12 T66 T199 T241 T258 T262 T276	N104 N135	WD domain, G-beta repeat: P114-D150, E156-D193, P72-D108, K30-G66 WD40 repeats: G111-D150, N153-D193, S69-D108, T31-G66 one copy of WD repeat (SMART): S110-D150, S69-D108, K155-D193 Trp-Asp (WD) repeat proteins BL00678: S139-W149 Trp-Asp (WD-40) repeats signature: Y84-F130, T127-S173 G-protein beta WD-40 repeat signature PR00320: V137-I151 Trp-Asp (WD) repeats signature: V137-I151	HMMER- TIGRFAM BLAST- PRODOM HMMER-PFAM HMMER-SMART HMMER-INCY BLIMPS- BLOCKS PROFILESKAN BLIMPS- PRINTS MOTIFS
14	7510809CD1	226	S42 S59 S110 S139 T87 T145 T159 T188 Y25 Y84			

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
15/2537210CB1/ 4344	1-655, 14-584, 63-4344, 65-572, 66-691, 170-751, 219-874, 244-960, 403-1010, 407-2745, 428-941, 562-726, 746-1023, 889-1048, 889-1259, 889-1417, 1027-1469, 1192-1536, 1547- 1822, 1547-2093, 1627-2152, 1688-2314, 1758-1819, 1986-2599, 1999-2521, 2005-2334, 2026- 2599, 2054-2724, 2214-2853, 2249-2828, 2358-2566, 2358-2581, 2414-3020, 2426-2622, 2478- 2787, 2506-2784, 2524-2767, 2524-3096, 2547-2815, 2709-2823, 2742-3263, 2841-3120, 2841- 3432, 2872-3122, 2932-3203, 2961-3413, 2972-3243, 2972-3520, 3057-3534, 3062-3747, 3075- 3270, 3087-3360, 3143-3627, 3178-3433, 3178-3455, 3181-3926, 3239-3910, 3278-3466, 3284- 3808, 3307-3943, 3367-3592, 3375-3575, 3662-3924, 3662-4119, 3670-3875, 3679-3950, 3681- 4223, 3687-4290, 3696-3911, 3751-4121, 3768-4159, 3813-4102, 3890-4147, 3912-4153, 3921- 4243, 3928-4210, 3952-4246, 3971-4195, 3974-4195, 3981-4344, 3988-4239, 4002-4264, 4009- 4249, 4055-4278, 4055-4317, 4087-4344, 4104-4251, 4128-4341
16/112535CB1/ 5621	1-908, 114-772, 167-1561, 202-544, 311-572, 810-1340, 854-1424, 871-1171, 981-1479, 1025- 1461, 1401-1630, 1423-1854, 1436-2068, 1635-2246, 1710-2122, 1773-2226, 1779-2319, 1843- 2139, 1843-2178, 1861-2313, 1906-2344, 1906-2356, 1920-2298, 2030-2701, 2062-2665, 2074- 2313, 2163-2666, 2184-2786, 2259-2695, 2262-2726, 2320-2539, 2320-2611, 2320-2862, 2320- 2898, 2320-2913, 2320-3003, 2329-2367, 2332-2367, 2334-2747, 2363-3046, 2392-2964, 2438- 2702, 2446-2581, 2446-2891, 2480-3174, 2491-2978, 2511-3197, 2546-2790, 2552-2761, 2582- 2867, 2584-2860, 2600-2878, 2613-3019, 2615-3244, 2712-3074, 2729-2932, 2730-2979, 2730- 3092, 2730-3095, 2730-3139, 2730-3182, 2730-3183, 2730-3190, 2730-3195, 2730-3198, 2730- 3223, 2730-3249, 2730-3254, 2730-3278, 2761-3338, 2790-2940, 2797-3334, 2801-3046, 2804- 3087, 2810-3345, 2818-3471, 2848-3380, 2849-3434, 2850-3493, 2921-3165, 2944-3451, 2995- 3373, 2995-3566, 2997-3235, 3008-3275, 3040-3729, 3048-3325, 3066-3554, 3066-3676, 3071- 3611, 3074-3397, 3080-3766, 3085-3893, 3095-3663, 3095-3759, 3097-3619, 3141-3719, 3158- 3813, 3161-3713, 3172-3703, 3172-3783, 3187-3853, 3210-3797, 3231-3470, 3235-3863, 3248- 3520, 3262-3567, 3268-3505, 3270-3626, 3294-3699, 3298-3901, 3319-3697, 3321-3986, 3323- 3594, 3347-4024, 3350-3985, 3356-3709, 3424-4149, 3439-3662, 3439-3669, 3439-3992, 3441- 3942, 3462-3776, 3462-3978, 3470-3713, 3485-4103, 3494-4007, 3501-4095, 3539-4173, 3541- 4032, 3753-4053, 3798-4261, 3834-4493, 3885-4170, 3961-4125, 3961-4216, 3980-4235, 3980- 4276, 3987-4552, 3997-4577, 4020-4143, 4030-4632, 4035-4637, 4056-4359, 4072-4530, 4098- 4301, 4107-4387, 4130-4775, 4142-4575, 4157-4806, 4178-4341, 4182-4621, 4183-4621, 4184- 4522, 4195-4981, 4216-4981, 4230-4499, 4248-4621, 4249-4799, 4258-4621, 4260-4486, 4261- 4930, 4285-4555, 4348-4584, 4348-4825, 4386-4673, 4405-4942, 4464-4774, 4485-4919,

Table 4 (cont.)

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
16	4506-4747, 4506-4992, 4510-4758, 4525-4981, 4534-4988, 4554-4841, 4554-4988, 4573-4979, 4580-4960, 4606-5166, 4674-4945, 4681-4967, 4692-4981, 4714-4981, 4714-4993, 4734-4988, 4734-4990, 4747-5317, 4863-5160, 4876-5114, 4886-5360, 4918-5563, 4935-5572, 4952-5186, 4977-5268, 4994-5562, 4994-5568, 4995-5252, 4995-5570, 5070-5568, 5077-5327, 5109-5592, 5150-5593, 5164-5595, 5200-5589, 5259-5621, 5302-5536, 5323-5508
17/72063274CB1/ 1581	1-252, 1-289, 1-698, 1-1551, 16-98, 117-437, 323-884, 324-925, 324-930, 365-610, 367-706, 587-1092, 605-1278, 605-1313, 605-1380, 605-1389, 605-1409, 605-1421, 605-1433, 644-1157, 663-1208, 672-1548, 676-1117, 692-1208, 704-1551, 719-1548, 719-1551, 738-1548, 740-1295, 752-1548, 755-1551, 762-945, 763-1551, 767-1140, 767-1155, 767-1188, 767-1211, 767-1354, 767-1402, 767-1420, 767-1557, 767-1574, 767-1575, 768-1094, 768-1433, 783-1551, 809-1551, 810-1551, 812-1506, 822-1465, 828-1513, 838-1578, 881-1548, 896-1496, 939-1496, 942-1512, 983-1245, 1034-1551, 1044-1551, 1071-1581, 1081-1581, 1111-1561, 1114-1564, 1134-1559, 1153-1558, 1178-1526, 1392-1498, 1498-1564
18/5013673CB1/ 3156	1-533, 1-619, 30-692, 41-111, 58-644, 234-833, 347-543, 418-959, 521-747, 525-1021, 542-668, 556-1279, 635-1309, 658-813, 747-808, 869-1428, 887-1495, 939-1476, 1089-1329, 1114-1770, 1115-1666, 1124-1804, 1133-1799, 1195-1409, 1275-1896, 1332-1995, 1352-1950, 1435-1823, 1765-2273, 1831-2505, 2083-2662, 2083-2683, 2083-2686, 2117-2710, 2160-2674, 2176-2862, 2190-2841, 2190-2845, 2404-2963, 2594-3156
19/5977982CB1/ 1567	1-354, 1-465, 1-485, 1-513, 1-628, 1-660, 1-663, 1-684, 1-689, 1-690, 1-825, 2-275, 2-283, 3-250, 3-296, 3-329, 5-492, 90-380, 211-877, 351-603, 362-664, 407-934, 447-552, 537-1259, 542-1390, 568-1340, 610-1291, 616-1567, 618-1158, 640-734, 664-1387
20/6880271CB1/ 2212	1-501, 1-532, 1-623, 5-532, 66-532, 212-1459, 392-843, 520-1161, 648-1354, 913-1420, 1120-1373, 1120-1434, 1120-1766, 1199-1750, 1227-1681, 1237-1629, 1240-1681, 1244-1617, 1244-1770, 1251-1840, 1290-1811, 1310-1681, 1336-1681, 1541-2212
21/2378756CB1/ 1751	1-257, 2-270, 2-466, 8-285, 14-594, 18-566, 18-645, 20-657, 23-265, 23-556, 23-595, 23-616, 67-307, 85-201, 270-889, 401-512, 408-640, 408-769, 408-902, 707-1428, 749-1424, 756-1424, 1043-1108, 1056-1506, 1119-1751, 1131-1606
22/1861527CB1/ 4189	1-229, 86-625, 413-795, 413-1001, 414-1001, 509-995, 509-1001, 542-1001, 553-915, 553-1000, 553-1001, 655-1001, 780-1001, 798-1001, 805-1314, 805-1315, 812-1389, 812-1508, 812-1519, 812-1527, 812-1534, 941-1728, 1014-1694, 1018-1694, 1021-1728, 1244-1493, 1262-1916, 1364-1989, 1422-1957, 1457-1873, 1463-2154, 1465-2197, 1467-2168, 1468-2148,

Table 4 (cont.)

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
22	1471-2267, 1478-2202, 1526-2154, 1528-2344, 1533-2452, 1553-2287, 1560-2241, 1575-2282, 1583-2263, 1643-2463, 1650-2414, 1689-2243, 1732-2265, 1732-2546, 1737-2389, 1738-2403, 1738-2464, 1841-2513, 1936-2646, 1991-2413, 2067-2676, 2070-2620, 2078-2752, 2079-2607, 2096-2801, 2105-2744, 2125-2711, 2139-2609, 2146-2680, 2173-2414, 2223-2842, 2236-2698, 2286-2929, 2524-3153, 2586-2798, 2619-3235, 2638-3098, 2681-2949, 2682-3219, 2729-3414, 2748-3358, 2764-3146, 2794-3355, 2798-3153, 2830-3455, 2831-3105, 2831-3513, 2856-3112, 2881-3357, 2912-3164, 2912-3407, 2912-3451, 2912-3470, 2912-3488, 2912-3542, 2912-3547, 2923-3551, 3028-3513, 3053-3523, 3062-3534, 3071-3331, 3071-3465, 3071-3509, 3071-3531, 3071-3578, 3071-3667, 3071-3719, 3071-3720, 3080-3783, 3094-3284, 3146-3812, 3151-3441, 3222-3462, 3222-3490, 3222-3611, 3297-3993, 3299-3543, 3302-3440, 3302-3846, 3342-3611, 3398-3901, 3411-3590, 3439-3705, 3454-3698, 3460-3754, 3488-3702, 3505-3716, 3505-3786, 3505-3793, 3505-4108, 3505-4161, 3506-3746, 3511-4119, 3528-3831, 3543-3772, 3586-4189, 3609-4170, 3659-3862, 3676-4169, 3819-4085, 3890-4114
23/2921356CB1/ 1679	1-693, 69-871, 194-734, 386-631, 386-817, 386-995, 395-802, 395-895, 395-946, 395-955, 404-1038, 420-1039, 502-1251, 585-1051, 598-899, 598-982, 600-972, 600-1164, 715-999, 715-1216, 715-1393, 715-1442, 715-1480, 715-1527, 715-1647, 715-1679, 726-1456, 867-1466, 985-1481, 1183-1459, 1249-1495, 1251-1625
24/7386170CB1/ 2077	1-463, 1-464, 42-404, 90-439, 93-341, 96-376, 139-374, 193-720, 461-1001, 462-736, 462- 791, 462-910, 462-958, 462-1023, 462-1024, 462-1112, 462-1142, 462-1144, 462-1146, 462- 1155, 462-1159, 462-1169, 462-1170, 462-1178, 462-1181, 462-1215, 462-1227, 462-1259, 462-1264, 462-1329, 462-1399, 463-1375, 473-1027, 476-926, 481-745, 494-1080, 496-1153, 497-681, 510-661, 512-911, 513-1058, 517-1001, 518-1379, 523-994, 524-948, 526-825, 527- 801, 527-1018, 530-1073, 534-1128, 536-1179, 538-1301, 546-805, 549-949, 556-1403, 560- 1073, 563-829, 564-864, 565-821, 565-858, 565-864, 567-673, 567-1361, 570-864, 573-864, 573-866, 573-871, 576-1070, 586-858, 587-864, 588-1066, 590-864, 597-723, 599-1019, 600- 1022, 612-864, 613-962, 626-895, 627-874, 627-1236, 630-1277, 633-934, 635-1365, 639- 1191, 639-1479, 643-1439, 657-1198, 658-1228, 660-921, 662-979, 664-1189, 669-1030, 877- 1441, 955-1535, 978-1421, 978-1426, 1001-1454, 1004-1413, 1027-1463, 1147-1584, 1250- 1515, 1255-1992, 1282-2028, 1287-2066, 1289-1517, 1305-1602, 1337-2077, 1345-1953, 1366- 2067, 1373-1620, 1377-2068

Table 4 (cont.)

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
25/7481206CB1/ 3498	1-206, 115-715, 122-671, 122-738, 122-880, 122-926, 126-317, 127-518, 135-565, 137-345, 137-358, 142-393, 201-854, 205-557, 205-626, 205-648, 205-705, 205-730, 205-753, 205-765, 205-782, 205-857, 205-900, 206-816, 209-606, 210-665, 210-802, 248-675, 342-642, 351-996, 377-1030, 383-744, 439-1030, 464-952, 465-1112, 520-1343, 539-1177, 593-1184, 643-1436, 746-1101, 824-1425, 843-1400, 878-1333, 916-1290, 925-1553, 940-1193, 975-1512, 1047- 1691, 1101-1711, 1189-1501, 1237-1766, 1250-1742, 1302-1660, 1312-1858, 1333-1531, 1364- 1921, 1401-1678, 1418-1930, 1440-1937, 1441-1837, 1450-1979, 1456-1587, 1456-2020, 1473- 2093, 1484-2110, 1487-1594, 1490-2041, 1509-2248, 1509-2418, 1519-2411, 1540-2052, 1547- 2150, 1552-2110, 1555-2126, 1565-2064, 1565-2250, 1587-2415, 1590-2107, 1603-2204, 1650- 2258, 1662-2231, 1731-2258, 1733-2308, 1774-2048, 1774-2242, 1795-2380, 1810-2457, 1814- 2573, 1818-2087, 1837-2482, 1842-2404, 1845-2340, 1846-2434, 1852-2387, 1852-2390, 1866- 2436, 1897-2501, 1920-2650, 1939-2521, 1942-2619, 1943-2536, 1949-2427, 1968-2496, 2097- 2717, 2104-2853, 2117-2901, 2139-2644, 2157-3070, 2159-2772, 2166-2748, 2183-2835, 2184- 2694, 2191-2787, 2221-2713, 2236-2816, 2259-2564, 2259-2638, 2259-2729, 2306-2906, 2309- 2766, 2341-2848, 2351-2982, 2358-2913, 2359-2933, 2415-2934, 2424-3256, 2436-3137, 2437- 2972, 2465-2984, 2467-2913, 2467-3077, 2490-3024, 2510-3203, 2531-3095, 2532-2924, 2535- 3121, 2551-3170, 2551-3191, 2555-3310, 2578-3165, 2579-3087, 2580-3210, 2581-2936, 2581- 3114, 2583-3420, 2583-3422, 2594-3100, 2597-3166, 2605-3026, 2608-3087, 2610-3051, 2623- 3193, 2625-2985, 2631-3285, 2645-3084, 2648-3152, 2649-3058, 2649-3095, 2649-3243, 2653- 3085, 2656-3206, 2658-3217, 2680-2959, 2680-3016, 2701-3320, 2702-2935, 2702-2970, 2702- 2973, 2702-2983, 2702-3239, 2702-3291, 2702-3359, 2709-2941, 2712-2973, 2715-3485, 2717- 2994, 2719-3290, 2727-2992, 2727-3017, 2729-3231, 2738-2934, 2738-3498, 2742-2978, 2743- 2905, 2746-3498, 2750-3220, 2754-3133, 2754-3314, 2759-2797, 2759-2965, 2759-3086, 2763- 3303, 2767-2986, 2768-2998, 2777-3056, 2780-3135, 2783-3070, 2784-3176, 2789-3376, 2796- 3302, 2811-3285, 2811-3498, 2812-3498, 2820-3424, 2828-3164, 2832-2894, 2837-3073, 2843- 3100, 2847-3449, 2852-3137, 2857-3099, 2857-3239, 2868-3207, 2872-3135, 2873-3474, 2878- 3375, 2884-3479, 2888-3248
26/7503117CB1/ 5540	1-908, 1-5540, 114-772, 311-572, 810-1340, 871-1171, 887-1359, 981-1479, 1401-1630, 1423- 1854, 1538-1858, 1615-1858, 1635-2246, 1710-2122, 1779-2319, 1843-2139, 1843-2178, 1889- 2349, 1892-2616, 1906-2344, 1906-2356, 1920-2298, 2090-2188, 2240-3008, 2320-2810, 2365- 2621, 2368-2883, 2410-2897, 2437-2711, 2465-2709, 2471-2680, 2501-2786, 2503-2779, 2519- 2797, 2534-3163, 2626-2731, 2648-2851, 2649-2898, 2649-3011, 2649-3014, 2649-3058,

Table 4 (cont.)

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
26	2649-3101, 2649-3102, 2649-3109, 2649-3114, 2649-3117, 2649-3142, 2649-3168, 2649-3173, 2649-3197, 2716-3253, 2723-3006, 2729-3264, 2737-3390, 2767-3299, 2768-3353, 2769-3412, 2840-3084, 2863-3370, 2871-3154, 2871-3200, 2871-3230, 2871-3328, 2914-3292, 2914-3485, 2916-3154, 2950-3225, 2959-3648, 2967-3244, 2985-3473, 2990-3530, 2993-3316, 2999-3685, 3004-3812, 3014-3582, 3014-3678, 3016-3538, 3017-3234, 3080-3632, 3091-3622, 3091-3702, 3106-3772, 3150-3389, 3167-3439, 3181-3486, 3187-3424, 3187-3654, 3189-3545, 3213-3618, 3214-3621, 3217-3820, 3242-3513, 3266-3943, 3269-3904, 3275-3628, 3310-3584, 3358-3581, 3358-3588, 3358-3911, 3360-3861, 3381-3897, 3389-3632, 3404-4022, 3420-4014, 3433-4007, 3460-3951, 3506-4098, 3575-4051, 3672-3972, 3717-4180, 3753-4412, 3804-4089, 3804-4157, 3880-4044, 3880-4135, 3899-4154, 3899-4195, 3939-4062, 3949-4551, 3954-4556, 3975-4278, 3991-4449, 4017-4220, 4026-4306, 4049-4694, 4083-4538, 4084-4538, 4093-4563, 4095-4539, 4101-4540, 4102-4540, 4114-4900, 4135-4900, 4149-4418, 4167-4540, 4168-4718, 4177-4540, 4179-4405, 4180-4849, 4193-4539, 4200-4540, 4204-4474, 4224-4538, 4267-4503, 4267-4744, 4294-4538, 4305-4592, 4383-4693, 4404-4838, 4411-4891, 4411-4900, 4413-4661, 4425-4666, 4425-4911, 4429-4677, 4442-4901, 4444-4900, 4453-4907, 4458-4900, 4473-4760, 4473-4905, 4492-4898, 4499-4879, 4515-4900, 4525-5085, 4541-4898, 4590-4869, 4593-4864, 4595-4776, 4600-4886, 4611-4898, 4624-4827, 4633-4900, 4633-4912, 4641-4898, 4653-4907, 4653-4909, 4666-5236, 4735-5217, 4782-5079, 4785-4900, 4795-5033, 4805-5279, 4837-5482, 4854-5491, 4871-5105, 4896-5187, 4913-5481, 4913-5487, 4914-5049, 4914-5171, 4914-5489, 4989-5487, 4996-5246, 5028-5511, 5057-5499, 5059-5491, 5083-5514, 5089-5494, 5104-5487, 5110-5498, 5110-5499, 5122-5493, 5132-5487, 5149-5487, 5161-5393, 5178-5453, 5178-5540, 5221-5455, 5242-5427, 5389-5540
27/7506911CB1/ 1785	1-650, 1-839, 1-850, 1-1768, 125-787, 125-956, 125-961, 125-985, 125-986, 182-931, 278- 579, 278-658, 280-652, 395-679, 395-896, 434-1346, 437-1346, 447-1346, 462-1346, 504- 1346, 519-1345, 562-1346, 594-1346, 633-1346, 639-1345, 685-1346, 723-1346, 863-1139, 869-1346, 1371-1785
28/7510809CB1/ 1615	1-257, 1-1615, 2-233, 2-270, 8-285, 18-566, 18-645, 20-657, 23-265, 23-595, 23-616, 67- 307, 85-201, 206-884, 206-899, 210-906, 210-911, 210-915, 270-889, 408-640, 408-769, 408- 902, 572-1128, 739-905, 983-1615

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID	Representative Library
15	2537210CB1	SINTFER02
16	112535CB1	BRSTNOT01
17	72063274CB1	OVARIN02
18	5013673CB1	BRAUTDR02
19	5977982CB1	MCLDXT02
20	6880271CB1	HNT2AGT01
21	2378756CB1	LIVRTUE01
22	1861527CB1	ENDVNOT01
23	2921356CB1	BRAIFER05
24	7386170CB1	BRAENOT04
25	7481206CB1	BRAINOT03
26	7503117CB1	BRSTNOT01
27	7506911CB1	LIVRNON08
28	7510809CB1	BRSTNOT05

Table 6

Library	Vector	Library Description
BRAENOT04	pINCY	Library was constructed using RNA isolated from inferior parietal cortex tissue removed from the brain of a 35-year-old Caucasian male who died from cardiac failure. Pathology indicated moderate leptomeningeal fibrosis and multiple microinfarctions of the cerebral neocortex. Patient history included dilated cardiomyopathy, congestive heart failure, cardiomegaly and an enlarged spleen and liver.
BRAIFER05	pINCY	Library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus who was stillborn with a hypoplastic left heart at 23 weeks' gestation.
BRAINOT03	PSPORT1	Library was constructed using RNA isolated from brain tissue removed from a 26-year-old Caucasian male during cranioplasty and excision of a cerebral meningeal lesion. Pathology for the associated tumor tissue indicated a grade 4 oligoastrocytoma in the right fronto-parietal part of the brain.
BRAUTDR02	PCDNA2.1	This random primed library was constructed using RNA isolated from pooled amygdala and entorhinal cortex tissue removed from a 55-year-old Caucasian female who died from cholangiocarcinoma. Pathology indicated mild meningeal fibrosis predominately over the convexities, scattered axonal spheroids in the white matter of the cingulate cortex and the thalamus, and a few scattered neurofibrillary tangles in the entorhinal cortex and the periaqueductal gray region. Pathology for the associated tumor tissue indicated well-differentiated cholangiocarcinoma of the liver with residual or relapsed tumor. Patient history included cholangiocarcinoma, post-operative Budd-Chiari syndrome, biliary ascites, hydrothorax, dehydration, malnutrition, oliguria and acute renal failure. Previous surgeries included cholecystectomy and resection of 85% of the liver.
BRSTNOT01	PBLUESCRIPT	Library was constructed using RNA isolated from the breast tissue of a 56-year-old Caucasian female who died in a motor vehicle accident.
BRSTNOT05	PSPORT1	Library was constructed using RNA isolated from breast tissue removed from a 58-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology for the associated tumor tissue indicated multicentric invasive grade 4 lobular carcinoma. Patient history included skin cancer, rheumatic heart disease, osteoarthritis, and tuberculosis. Family history included cerebrovascular and cardiovascular disease, breast and prostate cancer, and type I diabetes.

Table 6 (cont.)

Library	Vector	Library Description
ENDVNOT01	pINCY	Library was constructed using RNA isolated from untreated microvascular endothelial cell tissue removed from an 18-year-old Caucasian female.
HNT2AGT01	PBLUESCRIPT	Library was constructed at Stratagene (STR937233), using RNA isolated from the hNT2 cell line derived from a human teratocarcinoma that exhibited properties characteristic of a committed neuronal precursor. Cells were treated with retinoic acid for 5 weeks and with mitotic inhibitors for two weeks and allowed to mature for an additional 4 weeks in conditioned medium.
LIVRNON08	pINCY	This normalized library was constructed from 5.7 million independent clones from a pooled liver tissue library. Starting RNA was made from pooled liver tissue removed from a 4-year-old Hispanic male who died from anoxia and a 16 week female fetus who died after 16-weeks gestation from anencephaly. Serologies were positive for cytomegalovirus in the 4-year-old. Patient history included asthma in the 4-year-old. Family history included taking daily prenatal vitamins and mitral valve prolapse in the mother of the fetus. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
LIVRTUE01	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from liver tumor tissue removed from a 72-year-old Caucasian male during partial hepatectomy. Pathology indicated metastatic grade 2 (of 4) neuroendocrine carcinoma forming a mass. The patient presented with metastatic liver cancer. Patient history included benign hypertension, type I diabetes, prostatic hyperplasia, prostate cancer, alcohol abuse in remission, and tobacco abuse in remission. Previous surgeries included destruction of a pancreatic lesion, closed prostatic biopsy, transurethral prostatectomy, removal of bilateral testes and total splenectomy. Patient medications included Eulexin, Hytrin, Proscar, Ecotrin, and insulin. Family history included atherosclerotic coronary artery disease and acute myocardial infarction in the mother; atherosclerotic coronary artery disease and type II diabetes in the father.

Table 6 (cont.)

Library	Vector	Library Description
MCLDXTXT02	pINCY	Library was constructed using RNA isolated from treated umbilical cord blood dendritic cells removed from a male. The cells were treated with granulocyte/macrophage colony stimulating factor (GM-CSF), tumor necrosis factor alpha (TNF alpha), stem cell factor (SCF), phorbol myristate acetate (PMA), and ionomycin. The GM-CSF was added at time 0 at 100 ng/ml, the TNF alpha was added at time 0 at 2.5 ng/ml, the SCF was added at time 0 at 25 ng/ml. The PMA and ionomycin were added at 13 days for five hours. Incubation time was 13 days.
OVARDIN02	pINCY	This normalized ovarian tissue library was constructed from 5.76 million independent clones from an ovary library. Starting RNA was made from diseased ovarian tissue removed from a 39-year-old Caucasian female during total abdominal hysterectomy, bilateral salpingo-oophorectomy, dilation and curettage, partial colectomy, incidental appendectomy, and temporary colostomy. Pathology indicated the right and left adnexa, mesentery and muscularis propria of the sigmoid colon were extensively involved by endometriosis. Endometriosis also involved the anterior and posterior serosal surfaces of the uterus and the cul-de-sac. The endometrium was proliferative. Pathology for the associated tumor tissue indicated multiple (3 intramural, 1 subserosal) leiomyomata. The patient presented with abdominal pain and infertility. Patient history included scoliosis. Family history included hyperlipidemia, benign hypertension, atherosclerotic coronary artery disease, depressive disorder, brain cancer, and type II diabetes. The library was normalized in two rounds using conditions adapted from Soares et al., PNAS(1994) 91:9228 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48-hours/round) reannealing hybridization was used.
SINTFER02	pINCY	This random primed library was constructed using RNA isolated from small intestine tissue removed from a Caucasian male fetus who died from fetal demise.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value=1.0E-8 or less Full Length sequences: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value=1.06E-6 Assembled ESTs: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less Full Length sequences: fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value=1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM, INCY, SMART, and TIGRFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	PFAM, INCY, SMART, or TIGRFAM hits: Probability value=1.0E-3 or less Signal peptide hits: Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score=120 or greater; Match length=56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
27	7506911	2913521H1	SNP00043631	155	1210	C	C	T	H404	0.98	n/a	n/a	n/a
27	7506911	3281384H1	SNP00043630	89	1018	T	T	C	C340	n/a	n/a	n/a	n/a
27	7506911	3463658H1	SNP00043630	57	1017	T	T	C	I339	n/a	n/a	n/a	n/a
27	7506911	3907226H1	SNP00043630	120	1020	T	T	C	C340	n/a	n/a	n/a	n/a
27	7506911	4286341H1	SNP00043629	78	357	T	C	T	F119	n/a	n/a	n/a	n/a
27	7506911	4443694H1	SNP00043630	155	1016	C	T	C	T339	n/a	n/a	n/a	n/a
27	7506911	5882613H1	SNP00043629	79	357	C	C	T	F119	n/a	n/a	n/a	n/a
27	7506911	7207691H1	SNP00043630	474	1020	T	T	C	C340	n/a	n/a	n/a	n/a
27	7506911	7363877H1	SNP00043631	103	1212	C	C	T	D404	0.98	n/a	n/a	n/a
27	7506911	7637420H1	SNP00043629	484	359	C	C	T	P120	n/a	n/a	n/a	n/a

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:
 - a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-14,
 - b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3-8, SEQ ID NO:10, and SEQ ID NO:14,
 - c) a polypeptide comprising a naturally occurring amino acid sequence at least 91% identical to the amino acid sequence of SEQ ID NO:13,
 - d) a polypeptide comprising a naturally occurring amino acid sequence at least 99% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:9 and SEQ ID NO:11,
 - e) a polypeptide consisting essentially of a naturally occurring amino acid sequence at least 90% identical to the amino acid sequence of SEQ ID NO:12,
 - f) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and
 - g) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14.
2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-14.
3. An isolated polynucleotide encoding a polypeptide of claim 1.
4. An isolated polynucleotide encoding a polypeptide of claim 2.
5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28.
6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.
7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.
9. A method of producing a polypeptide of claim 1, the method comprising:
- 5 a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and
- b) recovering the polypeptide so expressed.
- 10 10. A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-14.
11. An isolated antibody which specifically binds to a polypeptide of claim 1.
- 15 12. An isolated polynucleotide selected from the group consisting of:
- a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28,
- b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of
- 20 SEQ ID NO:15-28,
- c) a polynucleotide complementary to a polynucleotide of a),
- d) a polynucleotide complementary to a polynucleotide of b), and
- e) an RNA equivalent of a)-d).
- 25 13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.
14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:
- 30 a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and
- 35 b) detecting the presence or absence of said hybridization complex, and, optionally, if

present, the amount thereof.

15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

5 16. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:
a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and
b) detecting the presence or absence of said amplified target polynucleotide or fragment
10 thereof, and, optionally, if present, the amount thereof.

17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

15 18. A composition of claim 17, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-14.

19. A method for treating a disease or condition associated with decreased expression of functional KPP, comprising administering to a patient in need of such treatment the composition of
20 claim 17.

20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:
a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
25 b) detecting agonist activity in the sample.

21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

30 22. A method for treating a disease or condition associated with decreased expression of functional KPP, comprising administering to a patient in need of such treatment a composition of claim 21.

23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of
35 claim 1, the method comprising:

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting antagonist activity in the sample.

24. A composition comprising an antagonist compound identified by a method of claim 23
5 and a pharmaceutically acceptable excipient.

25. A method for treating a disease or condition associated with overexpression of functional KPP, comprising administering to a patient in need of such treatment a composition of claim 24.

10 26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, the method comprising:

- a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and
- b) detecting binding of the polypeptide of claim 1 to the test compound, thereby
15 identifying a compound that specifically binds to the polypeptide of claim 1.

27. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

- a) combining the polypeptide of claim 1 with at least one test compound under
20 conditions permissive for the activity of the polypeptide of claim 1,
- b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and
- c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test
25 compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

28. A method of screening a compound for effectiveness in altering expression of a target
30 polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

- a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,
- b) detecting altered expression of the target polynucleotide, and
- c) comparing the expression of the target polynucleotide in the presence of varying
35

amounts of the compound and in the absence of the compound.

29. A method of assessing toxicity of a test compound, the method comprising:

- a) treating a biological sample containing nucleic acids with the test compound,
- 5 b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,
- 10 c) quantifying the amount of hybridization complex, and
- d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

15

30. A diagnostic test for a condition or disease associated with the expression of KPP in a biological sample, the method comprising:

- a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and
- 20 b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

31. The antibody of claim 11, wherein the antibody is:

- 25 a) a chimeric antibody,
- b) a single chain antibody,
- c) a Fab fragment,
- d) a F(ab')₂ fragment, or
- e) a humanized antibody.

30

32. A composition comprising an antibody of claim 11 and an acceptable excipient.

33. A method of diagnosing a condition or disease associated with the expression of KPP in a subject, comprising administering to said subject an effective amount of the composition of claim

35 32.

34. A composition of claim 32, wherein the antibody is labeled.

35. A method of diagnosing a condition or disease associated with the expression of KPP in a subject, comprising administering to said subject an effective amount of the composition of claim

5 34.

36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 11, the method comprising:

- 10 a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
- b) isolating antibodies from said animal, and
- c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which specifically binds to a polypeptide comprising an amino
- 15 acid sequence selected from the group consisting of SEQ ID NO:1-14.

37. A polyclonal antibody produced by a method of claim 36.

38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

20

39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:

- 25 a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
- b) isolating antibody producing cells from the animal,
- c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,
- d) culturing the hybridoma cells, and
- 30 e) isolating from the culture monoclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-14.

40. A monoclonal antibody produced by a method of claim 39.

35

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.

42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.

5

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.

44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-14 in a sample, the method comprising:

10

a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide comprising an amino acid sequence selected from the group consisting of

15

SEQ ID NO:1-14 in the sample.

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-14 from a sample, the method comprising:

a) incubating the antibody of claim 11 with a sample under conditions to allow specific

20

binding of the antibody and the polypeptide, and

b) separating the antibody from the sample and obtaining the purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-14.

25

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 13.

47. A method of generating an expression profile of a sample which contains polynucleotides, the method comprising:

30

a) labeling the polynucleotides of the sample,

b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and

c) quantifying the expression of the polynucleotides in the sample.

35

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.

52. An array of claim 48, which is a microarray.

53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.

55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical location have the same sequence, and each distinct physical location on the substrate contains nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.
60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.
- 5 61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.
62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.
63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.
- 10 64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.
65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.
- 15 66. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.
67. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.
68. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.
- 20 69. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.
70. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
NO:15.
- 25 71. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
NO:16.
72. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
30 NO:17.
73. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
NO:18.
- 35 74. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID

NO:19.

75. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
NO:20.

5

76. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
NO:21.

77. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
10 NO:22.

78. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
NO:23.

15 79. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
NO:24.

80. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
NO:25.

20

81. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
NO:26.

82. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
25 NO:27.

83. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
NO:28.

<110> INCYTE GENOMICS, INC.

YUE, Henry
 LU, Dyung Aina M.
 AZIMZAI, Yalda
 DING, Li
 LEE, Ernestine A.
 HAFALIA, April J.A.
 BECHA, Shanya D.
 TANG, Y. Tom
 LAL, Preeti G.
 GRIFFIN, Jennifer A.
 GURURAJAN, Rajagopal
 RAMKUMAR, Jayalaxmi
 ELLIOTT, Vicki S.
 ARVIZU, Chandra
 LUO, Wen
 SWARNAKAR, Anita
 DUGGAN, Brendan M.
 TRAN, Uyen K.
 WALIA, Narinder K.
 GANDHI, Ameena R.
 YAO, Monique G.
 KHAN, Farrah A.
 BAUGHN, Mariah R.
 BOROWSKY, Mark L.
 ZEBARJADIAN, Yeganeh
 RICHARDSON, Thomas W.
 MARQUIS, Joseph P.
 CHIEN, David
 JIN, Pei

<120> KINASES AND PHOSPHATASES

<130> PF-0995 PCT

<140> To Be Assigned

<141> Herewith

<150> 60/293,665; 60/298,712; 60/303,418; 60/306,967; 60/308,183;
 60/343,007; 60/357,675; 60/376,988

<151> 2001-05-24; 2001-06-15; 2001-07-06; 2001-07-19; 2001-07-27;
 2001-12-19; 2002-02-15; 2002-04-30

<160> 28

<170> PERL Program

<210> 1

<211> 856

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 2537210CD1

<400> 1

Met	Gly	Thr	Thr	Ala	Ser	Thr	Ala	Gln	Gln	Thr	Val	Ser	Ala	Gly
1				5				10						15
Thr	Pro	Phe	Glu	Gly	Leu	Gln	Gly	Ser	Gly	Thr	Met	Asp	Ser	Arg
			20					25						30
His	Ser	Val	Ser	Ile	His	Ser	Phe	Gln	Ser	Thr	Ser	Leu	His	Asn
			35					40						45

Ser	Lys	Ala	Lys	Ser	Ile	Ile	Pro	Asn	Lys	Val	Ala	Pro	Val	Val
				50					55					60
Ile	Thr	Tyr	Asn	Cys	Lys	Glu	Glu	Phe	Gln	Ile	His	Asp	Glu	Leu
				65					70					75
Leu	Lys	Ala	His	Tyr	Thr	Leu	Gly	Arg	Leu	Ser	Asp	Asn	Thr	Pro
				80					85					90
Glu	His	Tyr	Leu	Val	Gln	Gly	Arg	Tyr	Phe	Leu	Val	Arg	Asp	Val
				95					100					105
Thr	Glu	Lys	Met	Asp	Val	Leu	Gly	Thr	Val	Gly	Ser	Cys	Gly	Ala
				110					115					120
Pro	Asn	Phe	Arg	Gln	Val	Gln	Gly	Gly	Leu	Thr	Val	Phe	Gly	Met
				125					130					135
Gly	Gln	Pro	Ser	Leu	Ser	Gly	Phe	Arg	Arg	Val	Leu	Gln	Lys	Leu
				140					145					150
Gln	Lys	Asp	Gly	His	Arg	Glu	Cys	Val	Ile	Phe	Cys	Val	Arg	Glu
				155					160					165
Glu	Pro	Val	Leu	Phe	Leu	Arg	Ala	Asp	Glu	Asp	Phe	Val	Ser	Tyr
				170					175					180
Thr	Pro	Arg	Asp	Lys	Gln	Asn	Leu	His	Glu	Asn	Leu	Gln	Gly	Leu
				185					190					195
Gly	Pro	Gly	Val	Arg	Val	Glu	Ser	Leu	Glu	Leu	Ala	Ile	Arg	Lys
				200					205					210
Glu	Ile	His	Asp	Phe	Ala	Gln	Leu	Ser	Glu	Asn	Thr	Tyr	His	Val
				215					220					225
Tyr	His	Asn	Thr	Glu	Asp	Leu	Trp	Gly	Glu	Pro	His	Ala	Val	Ala
				230					235					240
Ile	His	Gly	Glu	Asp	Asp	Leu	His	Val	Thr	Glu	Glu	Val	Tyr	Lys
				245					250					255
Arg	Pro	Leu	Phe	Leu	Gln	Pro	Thr	Tyr	Arg	Tyr	His	Arg	Leu	Pro
				260					265					270
Leu	Pro	Glu	Gln	Gly	Ser	Pro	Leu	Glu	Ala	Gln	Leu	Asp	Ala	Phe
				275					280					285
Val	Ser	Val	Leu	Arg	Glu	Thr	Pro	Ser	Leu	Leu	Gln	Leu	Arg	Asp
				290					295					300
Ala	His	Gly	Pro	Pro	Pro	Ala	Leu	Val	Phe	Ser	Cys	Gln	Met	Gly
				305					310					315
Val	Gly	Arg	Thr	Asn	Leu	Gly	Met	Val	Leu	Gly	Thr	Leu	Ile	Leu
				320					325					330
Leu	His	Arg	Ser	Gly	Thr	Thr	Ser	Gln	Pro	Glu	Ala	Ala	Pro	Thr
				335					340					345
Gln	Ala	Lys	Pro	Leu	Pro	Met	Glu	Gln	Phe	Gln	Val	Ile	Gln	Ser
				350					355					360
Phe	Leu	Arg	Met	Val	Pro	Gln	Gly	Arg	Arg	Met	Val	Glu	Glu	Val
				365					370					375
Asp	Arg	Ala	Ile	Thr	Ala	Cys	Ala	Glu	Leu	His	Asp	Leu	Lys	Glu
				380					385					390
Val	Val	Leu	Glu	Asn	Gln	Lys	Lys	Leu	Glu	Gly	Ile	Arg	Pro	Glu
				395					400					405
Ser	Pro	Ala	Gln	Gly	Ser	Gly	Ser	Arg	His	Ser	Val	Trp	Gln	Arg
				410					415					420
Ala	Leu	Trp	Ser	Leu	Glu	Arg	Tyr	Phe	Tyr	Leu	Ile	Leu	Phe	Asn
				425					430					435
Tyr	Tyr	Leu	His	Glu	Gln	Tyr	Pro	Leu	Ala	Phe	Ala	Leu	Ser	Phe
				440					445					450
Ser	Arg	Trp	Leu	Cys	Ala	His	Pro	Glu	Leu	Tyr	Arg	Leu	Pro	Val
				455					460					465
Thr	Leu	Ser	Ser	Ala	Gly	Pro	Val	Ala	Pro	Arg	Asp	Leu	Ile	Ala
				470					475					480
Arg	Gly	Ser	Leu	Arg	Glu	Asp	Asp	Leu	Val	Ser	Pro	Asp	Ala	Leu
				485					490					495
Ser	Thr	Val	Arg	Glu	Met	Asp	Val	Ala	Asn	Phe	Arg	Arg	Val	Pro
				500					505					510
Arg	Met	Pro	Ile	Tyr	Gly	Thr	Ala	Gln	Pro	Ser	Ala	Lys	Ala	Leu

	515		520		525
Gly Ser Ile Leu	Ala Tyr Leu Thr Asp	Ala Lys Arg Arg Leu Arg			
	530		535		540
Lys Val Val Trp	Val Ser Leu Arg Glu	Glu Ala Val Leu Glu Cys			
	545		550		555
Asp Gly His Thr	Tyr Ser Leu Arg Trp	Pro Gly Pro Pro Val Ala			
	560		565		570
Pro Asp Gln Leu	Glu Thr Leu Glu Ala	Gln Leu Lys Ala His Leu			
	575		580		585
Ser Glu Pro Pro	Pro Gly Lys Glu Gly	Pro Leu Thr Tyr Arg Phe			
	590		595		600
Gln Thr Cys Leu	Thr Met Gln Glu Val	Phe Ser Gln His Arg Arg			
	605		610		615
Ala Cys Pro Gly	Leu Thr Tyr His Arg	Ile Pro Met Pro Asp Phe			
	620		625		630
Cys Ala Pro Arg	Glu Glu Asp Phe Asp	Gln Leu Leu Glu Ala Leu			
	635		640		645
Arg Ala Ala Leu	Ser Lys Asp Pro Gly	Thr Gly Phe Val Phe Ser			
	650		655		660
Cys Leu Ser Gly	Gln Gly Arg Thr Thr	Thr Ala Met Val Val Ala			
	665		670		675
Val Leu Ala Phe	Trp His Ile Gln Gly	Phe Pro Glu Val Gly Glu			
	680		685		690
Glu Glu Leu Val	Ser Val Pro Asp Ala	Lys Phe Thr Lys Gly Glu			
	695		700		705
Phe Gln Val Val	Met Lys Val Val Gln	Leu Leu Pro Asp Gly His			
	710		715		720
Arg Val Lys Lys	Glu Val Asp Ala Ala	Leu Asp Thr Val Ser Glu			
	725		730		735
Thr Met Thr Pro	Met His Tyr His Leu	Arg Glu Ile Ile Ile Cys			
	740		745		750
Thr Tyr Arg Gln	Ala Lys Ala Ala Lys	Glu Ala Gln Glu Met Arg			
	755		760		765
Arg Leu Gln Leu	Arg Ser Leu Gln Tyr	Leu Glu Arg Tyr Val Cys			
	770		775		780
Leu Ile Leu Phe	Asn Ala Tyr Leu His	Leu Glu Lys Ala Asp Ser			
	785		790		795
Trp Gln Arg Pro	Phe Ser Thr Trp Met	Gln Glu Val Ala Ser Lys			
	800		805		810
Ala Gly Ile Tyr	Glu Ile Leu Asn Glu	Leu Gly Phe Pro Glu Leu			
	815		820		825
Glu Ser Gly Glu	Asp Gln Pro Phe Ser	Arg Leu Arg Tyr Arg Trp			
	830		835		840
Gln Glu Gln Ser	Cys Ser Leu Glu Pro	Ser Ala Pro Glu Asp Leu			
	845		850		855
Leu					

<210> 2

<211> 1036

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 112535CD1

<400> 2

Met Glu Val Val	Gly Asp Phe Glu Tyr	Ser Lys Arg Asp Leu Val
1	5	10
Gly His Gly Ala	Phe Ala Val Val Phe	Arg Gly Arg His Arg Gln
	20	25
Lys Thr Asp Trp	Glu Val Ala Ile Lys	Ser Ile Asn Lys Lys Asn

				35					40					45
Leu	Ser	Lys	Ser	Gln	Ile	Leu	Leu	Gly	Lys	Glu	Ile	Lys	Ile	Leu
				50					55					60
Lys	Glu	Leu	Gln	His	Glu	Asn	Ile	Val	Ala	Leu	Tyr	Asp	Val	Gln
				65					70					75
Glu	Leu	Pro	Asn	Ser	Val	Phe	Leu	Val	Met	Glu	Tyr	Cys	Asn	Gly
				80					85					90
Gly	Asp	Leu	Ala	Asp	Tyr	Leu	Gln	Ala	Lys	Gly	Thr	Leu	Ser	Glu
				95					100					105
Asp	Thr	Ile	Arg	Val	Phe	Leu	His	Gln	Ile	Ala	Ala	Ala	Met	Arg
				110					115					120
Ile	Leu	His	Ser	Lys	Gly	Ile	Ile	His	Arg	Asp	Leu	Lys	Pro	Gln
				125					130					135
Asn	Ile	Leu	Leu	Ser	Tyr	Ala	Asn	Arg	Arg	Lys	Ser	Ser	Val	Ser
				140					145					150
Gly	Ile	Arg	Ile	Lys	Ile	Ala	Asp	Phe	Gly	Phe	Ala	Arg	Tyr	Leu
				155					160					165
His	Ser	Asn	Met	Met	Ala	Ala	Thr	Leu	Cys	Gly	Ser	Pro	Met	Tyr
				170					175					180
Met	Ala	Pro	Glu	Val	Ile	Met	Ser	Gln	His	Tyr	Asp	Ala	Lys	Ala
				185					190					195
Asp	Leu	Trp	Ser	Ile	Gly	Thr	Val	Ile	Tyr	Gln	Cys	Leu	Val	Gly
				200					205					210
Lys	Pro	Pro	Phe	Gln	Ala	Asn	Ser	Pro	Gln	Asp	Leu	Arg	Met	Phe
				215					220					225
Tyr	Glu	Lys	Asn	Arg	Ser	Leu	Met	Pro	Ser	Ile	Pro	Arg	Glu	Thr
				230					235					240
Ser	Pro	Tyr	Leu	Ala	Asn	Leu	Leu	Leu	Gly	Leu	Leu	Gln	Arg	Asn
				245					250					255
Gln	Lys	Asp	Arg	Met	Asp	Phe	Glu	Ala	Phe	Phe	Ser	His	Pro	Phe
				260					265					270
Leu	Glu	Gln	Gly	Pro	Val	Lys	Lys	Ser	Cys	Pro	Val	Pro	Val	Pro
				275					280					285
Met	Tyr	Ser	Gly	Ser	Val	Ser	Gly	Ser	Ser	Cys	Gly	Ser	Ser	Pro
				290					295					300
Ser	Cys	Arg	Phe	Ala	Ser	Pro	Pro	Ser	Leu	Pro	Asp	Met	Gln	His
				305					310					315
Ile	Gln	Glu	Glu	Asn	Leu	Ser	Ser	Pro	Pro	Leu	Gly	Pro	Pro	Asn
				320					325					330
Tyr	Leu	Gln	Val	Ser	Lys	Asp	Ser	Ala	Ser	Thr	Ser	Ser	Lys	Asn
				335					340					345
Ser	Ser	Cys	Asp	Thr	Asp	Asp	Phe	Val	Leu	Val	Pro	His	Asn	Ile
				350					355					360
Ser	Ser	Asp	His	Ser	Cys	Asp	Met	Pro	Met	Gly	Thr	Ala	Gly	Arg
				365					370					375
Arg	Ala	Ser	Asn	Glu	Phe	Leu	Val	Cys	Gly	Gly	Gln	Cys	Gln	Pro
				380					385					390
Thr	Val	Ser	Pro	His	Ser	Glu	Thr	Ala	Pro	Ile	Pro	Val	Pro	Thr
				395					400					405
Gln	Ile	Arg	Asn	Tyr	Gln	Arg	Ile	Glu	Gln	Asn	Leu	Thr	Ser	Thr
				410					415					420
Ala	Ser	Ser	Gly	Thr	Asn	Val	His	Gly	Ser	Pro	Arg	Ser	Ala	Val
				425					430					435
Val	Arg	Arg	Ser	Asn	Thr	Ser	Pro	Met	Gly	Phe	Leu	Arg	Pro	Gly
				440					445					450
Ser	Cys	Ser	Pro	Val	Pro	Ala	Asp	Thr	Ala	Gln	Thr	Val	Gly	Arg
				455					460					465
Arg	Leu	Ser	Thr	Gly	Ser	Ser	Arg	Pro	Tyr	Ser	Pro	Ser	Pro	Leu
				470					475					480
Val	Gly	Thr	Ile	Pro	Glu	Gln	Phe	Ser	Gln	Cys	Cys	Cys	Gly	His
				485					490					495
Pro	Gln	Gly	His	Asp	Ser	Arg	Ser	Arg	Asn	Ser	Ser	Gly	Ser	Pro
				500					505					510

Val	Pro	Gln	Ala	Gln	Ser	Pro	Gln	Ser	Leu	Leu	Ser	Gly	Ala	Arg
				515					520					525
Leu	Gln	Ser	Ala	Pro	Thr	Leu	Thr	Asp	Ile	Tyr	Gln	Asn	Lys	Gln
				530					535					540
Lys	Leu	Arg	Lys	Gln	His	Ser	Asp	Pro	Val	Cys	Pro	Ser	His	Thr
				545					550					555
Gly	Ala	Gly	Tyr	Ser	Tyr	Ser	Pro	Gln	Pro	Ser	Arg	Pro	Gly	Ser
				560					565					570
Leu	Gly	Thr	Ser	Pro	Thr	Lys	His	Leu	Gly	Ser	Ser	Pro	Arg	Ser
				575					580					585
Ser	Asp	Trp	Phe	Phe	Lys	Thr	Pro	Leu	Pro	Thr	Ile	Ile	Gly	Ser
				590					595					600
Pro	Thr	Lys	Thr	Thr	Ala	Pro	Phe	Lys	Ile	Pro	Lys	Thr	Gln	Ala
				605					610					615
Ser	Ser	Asn	Leu	Leu	Ala	Leu	Val	Thr	Arg	His	Gly	Pro	Ala	Glu
				620					625					630
Glu	Gln	Ser	Lys	Asp	Gly	Asn	Glu	Pro	Arg	Glu	Cys	Ala	His	Cys
				635					640					645
Leu	Leu	Val	Gln	Gly	Ser	Glu	Arg	Gln	Arg	Ala	Glu	Gln	Gln	Ser
				650					655					660
Lys	Ala	Val	Phe	Gly	Arg	Ser	Val	Ser	Thr	Gly	Lys	Leu	Ser	Asp
				665					670					675
Gln	Gln	Gly	Lys	Thr	Pro	Ile	Cys	Arg	His	Gln	Gly	Ser	Thr	Asp
				680					685					690
Ser	Leu	Asn	Thr	Glu	Arg	Pro	Met	Asp	Ile	Ala	Pro	Ala	Gly	Ala
				695					700					705
Cys	Gly	Gly	Val	Leu	Ala	Pro	Pro	Ala	Gly	Thr	Ala	Ala	Ser	Ser
				710					715					720
Lys	Ala	Val	Leu	Phe	Thr	Val	Gly	Ser	Pro	Pro	His	Ser	Ala	Ala
				725					730					735
Ala	Pro	Thr	Cys	Thr	His	Met	Phe	Leu	Arg	Thr	Arg	Thr	Thr	Ser
				740					745					750
Val	Gly	Pro	Ser	Asn	Ser	Gly	Gly	Ser	Leu	Cys	Ala	Met	Ser	Gly
				755					760					765
Arg	Val	Cys	Val	Gly	Ser	Pro	Pro	Gly	Pro	Gly	Phe	Gly	Ser	Ser
				770					775					780
Pro	Pro	Gly	Ala	Glu	Ala	Ala	Pro	Ser	Leu	Arg	Tyr	Val	Pro	Tyr
				785					790					795
Gly	Ala	Ser	Pro	Pro	Ser	Leu	Glu	Gly	Leu	Ile	Thr	Phe	Glu	Ala
				800					805					810
Pro	Glu	Leu	Pro	Glu	Glu	Thr	Leu	Met	Glu	Arg	Glu	His	Thr	Asp
				815					820					825
Thr	Leu	Arg	His	Leu	Asn	Val	Met	Leu	Met	Phe	Thr	Glu	Cys	Val
				830					835					840
Leu	Asp	Leu	Thr	Ala	Met	Arg	Gly	Gly	Asn	Pro	Glu	Leu	Cys	Thr
				845					850					855
Ser	Ala	Val	Ser	Leu	Tyr	Gln	Ile	Gln	Glu	Ser	Val	Val	Val	Asp
				860					865					870
Gln	Ile	Ser	Gln	Leu	Ser	Lys	Asp	Trp	Gly	Arg	Val	Glu	Gln	Leu
				875					880					885
Val	Leu	Tyr	Met	Lys	Ala	Ala	Gln	Leu	Leu	Ala	Ala	Ser	Leu	His
				890					895					900
Leu	Ala	Lys	Ala	Gln	Ile	Lys	Ser	Gly	Lys	Leu	Ser	Pro	Ser	Thr
				905					910					915
Ala	Val	Lys	Gln	Val	Val	Lys	Asn	Leu	Asn	Glu	Arg	Tyr	Lys	Phe
				920					925					930
Cys	Ile	Thr	Met	Cys	Lys	Lys	Leu	Thr	Glu	Lys	Leu	Asn	Arg	Phe
				935					940					945
Phe	Ser	Asp	Lys	Gln	Arg	Phe	Ile	Asp	Glu	Ile	Asn	Ser	Val	Thr
				950					955					960
Ala	Glu	Lys	Leu	Ile	Tyr	Asn	Cys	Ala	Val	Glu	Met	Val	Gln	Ser
				965					970					975
Ala	Ala	Leu	Asp	Glu	Met	Phe	Gln	Gln	Thr	Glu	Asp	Ile	Val	Tyr

	980		985		990
Arg Tyr His Lys	Ala Ala Leu Leu Leu	Glu Gly Leu Ser Arg	Ile		
	995		1000		1005
Leu Gln Asp Pro	Ala Asp Ile Glu Asn Val	His Lys Tyr Lys Cys			
	1010		1015		1020
Ser Ile Glu Arg	Arg Leu Ser Ala Leu Cys	His Ser Thr Ala Thr			
	1025		1030		1035
Val					

<210> 3

<211> 479

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 72063274CD1

<400> 3

Met Asp Ala Thr	Ile Ala Pro His Arg	Ile Pro Pro Glu Met	Pro
1	5	10	15
Gln Tyr Gly Glu	Asn His Ile Phe Glu	Leu Met Gln Asn	Met
	20	25	30
Leu Glu Gln Leu	Leu Ile His Gln Pro	Glu Asp Pro Ile	Pro Phe
	35	40	45
Met Ile Gln His	Leu His Arg Asp Asn	Asp Asn Val Pro	Arg Ile
	50	55	60
Val Ile Leu Gly	Pro Pro Ala Ser Gly	Lys Thr Thr Ile	Ala Met
	65	70	75
Trp Leu Cys Lys	His Leu Asn Ser Ser	Leu Leu Thr Leu	Glu Asn
	80	85	90
Leu Ile Leu Asn	Glu Phe Ser Tyr Thr	Ala Thr Glu Ala	Arg Arg
	95	100	105
Leu Tyr Leu Gln	Arg Lys Thr Val Pro	Ser Ala Leu Leu	Val Gln
	110	115	120
Leu Ile Gln Glu	Arg Leu Ala Glu Glu	Asp Cys Ile Lys	Gln Gly
	125	130	135
Trp Ile Leu Asp	Gly Ile Pro Glu Thr	Arg Glu Gln Ala	Leu Arg
	140	145	150
Ile Gln Thr Leu	Gly Ile Thr Pro Arg	His Val Ile Val	Leu Ser
	155	160	165
Ala Pro Asp Thr	Val Leu Ile Glu Arg	Asn Leu Gly Lys	Arg Ile
	170	175	180
Asp Pro Gln Thr	Gly Glu Ile Tyr His	Thr Thr Phe Asp	Trp Pro
	185	190	195
Pro Glu Ser Glu	Ile Gln Asn Arg Leu	Met Val Pro Glu	Asp Ile
	200	205	210
Ser Glu Leu Glu	Thr Ala Gln Lys Leu	Leu Glu Tyr His	Arg Asn
	215	220	225
Ile Val Arg Val	Ile Pro Ser Tyr Pro	Lys Ile Leu Lys	Val Ile
	230	235	240
Ser Ala Asp Gln	Pro Cys Val Asp Val	Phe Tyr Gln Ala	Leu Thr
	245	250	255
Tyr Val Gln Ser	Asn His Arg Thr Asn	Ala Pro Phe Thr	Pro Arg
	260	265	270
Val Leu Leu Leu	Gly Pro Val Gly Ser	Gly Lys Ser Leu	Gln Ala
	275	280	285
Ala Leu Leu Ala	Gln Lys Tyr Arg Leu	Val Asn Val Cys	Cys Gly
	290	295	300
Gln Leu Leu Lys	Glu Ala Val Ala Asp	Arg Thr Thr Phe	Gly Glu
	305	310	315
Leu Ile Gln Pro	Phe Phe Glu Lys Glu	Met Ala Val Pro	Asp Ser

Leu	Leu	Met	Lys	320	Val	Leu	Ser	Gln	Arg	325	Leu	Asp	Gln	Gln	Asp	330	Cys
				335						340						345	
Ile	Gln	Lys	Gly	350	Trp	Val	Leu	His	Gly	355	Val	Pro	Arg	Asp	Leu	360	Asp
Gln	Ala	His	Leu	365	Leu	Asn	Arg	Leu	Gly	370	Tyr	Asn	Pro	Asn	Arg	375	Val
Phe	Phe	Leu	Asn	380	Val	Pro	Phe	Asp	Ser	385	Ile	Met	Glu	Arg	Leu	390	Thr
Leu	Arg	Arg	Ile	395	Asp	Pro	Val	Thr	Gly	400	Glu	Arg	Tyr	His	Leu	405	Met
Tyr	Lys	Pro	Pro	410	Pro	Thr	Met	Glu	Ile	415	Gln	Ala	Arg	Leu	Leu	420	Gln
Asn	Pro	Lys	Asp	425	Ala	Glu	Glu	Gln	Val	430	Lys	Leu	Lys	Met	Asp	435	Leu
Phe	Tyr	Arg	Asn	440	Ser	Ala	Asp	Leu	Glu	445	Gln	Leu	Tyr	Gly	Ser	450	Ala
Ile	Thr	Leu	Asn	455	Gly	Asp	Gln	Asp	Pro	460	Tyr	Thr	Val	Phe	Glu	465	Tyr
Ile	Glu	Ser	Gly	470	Ile	Ile	Asn	Pro	Leu	475	Pro	Lys	Lys	Ile	Pro		

<210> 4

<211> 760

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 5013673CD1

<400> 4

Met	Val	Leu	Glu	Gln	Tyr	Val	Val	Val	Ala	Asn	Tyr	Gln	Lys	Gln			
1				5					10					15			
Glu	Ser	Ser	Glu	Ile	Ser	Leu	Ser	Val	Gly	Gln	Val	Val	Asp	Ile			
				20					25					30			
Ile	Glu	Lys	Asn	Glu	Ser	Gly	Trp	Trp	Phe	Val	Ser	Thr	Ala	Glu			
				35					40					45			
Glu	Gln	Gly	Trp	Val	Pro	Ala	Thr	Cys	Leu	Glu	Gly	Gln	Asp	Gly			
				50					55					60			
Val	Gln	Asp	Glu	Phe	Ser	Leu	Gln	Pro	Glu	Glu	Glu	Glu	Lys	Tyr			
				65					70					75			
Thr	Val	Ile	Tyr	Pro	Tyr	Thr	Ala	Arg	Asp	Gln	Asp	Glu	Met	Asn			
				80					85					90			
Leu	Glu	Arg	Gly	Ala	Val	Val	Glu	Val	Ile	Gln	Lys	Asn	Leu	Glu			
				95					100					105			
Gly	Trp	Trp	Lys	Ile	Arg	Tyr	Gln	Gly	Lys	Glu	Gly	Trp	Ala	Pro			
				110					115					120			
Ala	Ser	Tyr	Leu	Lys	Lys	Asn	Ser	Gly	Glu	Pro	Leu	Pro	Pro	Lys			
				125					130					135			
Pro	Gly	Pro	Gly	Ser	Pro	Ser	His	Pro	Gly	Ala	Leu	Asp	Leu	Asp			
				140					145					150			
Gly	Val	Ser	Arg	Gln	Gln	Asn	Ala	Val	Gly	Arg	Glu	Lys	Glu	Leu			
				155					160					165			
Leu	Ser	Ser	Gln	Arg	Asp	Gly	Arg	Phe	Glu	Gly	Arg	Pro	Val	Pro			
				170					175					180			
Asp	Gly	Asp	Ala	Lys	Gln	Arg	Ser	Pro	Lys	Met	Arg	Gln	Arg	Pro			
				185					190					195			
Pro	Pro	Arg	Arg	Asp	Met	Thr	Ile	Pro	Arg	Gly	Leu	Asn	Leu	Pro			
				200					205					210			
Lys	Pro	Pro	Ile	Pro	Pro	Gln	Val	Glu	Glu	Glu	Tyr	Tyr	Thr	Ile			
				215					220					225			
Ala	Glu	Phe	Gln	Thr	Thr	Ile	Pro	Asp	Gly	Ile	Ser	Phe	Gln	Ala			

				230					235					240
Gly	Leu	Lys	Val	Glu	Val	Ile	Glu	Lys	Asn	Leu	Ser	Gly	Trp	Trp
				245					250					255
Tyr	Ile	Gln	Ile	Glu	Asp	Lys	Glu	Gly	Trp	Ala	Pro	Ala	Thr	Phe
				260					265					270
Ile	Asp	Lys	Tyr	Lys	Lys	Thr	Ser	Asn	Ala	Ser	Arg	Pro	Asn	Phe
				275					280					285
Leu	Ala	Pro	Leu	Pro	His	Glu	Val	Thr	Gln	Leu	Arg	Leu	Gly	Glu
				290					295					300
Ala	Ala	Ala	Leu	Glu	Asn	Asn	Thr	Gly	Ser	Glu	Ala	Thr	Gly	Pro
				305					310					315
Ser	Arg	Pro	Leu	Pro	Asp	Ala	Pro	His	Gly	Val	Met	Asp	Ser	Gly
				320					325					330
Leu	Pro	Trp	Ser	Lys	Asp	Trp	Lys	Gly	Ser	Lys	Asp	Val	Leu	Arg
				335					340					345
Lys	Ala	Ser	Ser	Asp	Met	Ser	Ala	Ser	Ala	Gly	Tyr	Glu	Glu	Ile
				350					355					360
Ser	Asp	Pro	Asp	Met	Glu	Glu	Lys	Pro	Ser	Leu	Pro	Pro	Arg	Lys
				365					370					375
Glu	Ser	Ile	Ile	Lys	Ser	Glu	Gly	Glu	Leu	Leu	Glu	Arg	Glu	Arg
				380					385					390
Glu	Arg	Gln	Arg	Thr	Glu	Gln	Leu	Arg	Gly	Pro	Thr	Pro	Lys	Pro
				395					400					405
Pro	Gly	Val	Ile	Leu	Pro	Met	Met	Pro	Ala	Lys	His	Ile	Pro	Pro
				410					415					420
Ala	Arg	Asp	Ser	Arg	Arg	Pro	Glu	Pro	Lys	Pro	Asp	Lys	Ser	Arg
				425					430					435
Leu	Phe	Gln	Leu	Lys	Asn	Asp	Met	Gly	Leu	Glu	Cys	Gly	His	Lys
				440					445					450
Val	Leu	Ala	Lys	Glu	Val	Lys	Lys	Pro	Asn	Leu	Arg	Pro	Ile	Ser
				455					460					465
Lys	Ser	Lys	Thr	Asp	Leu	Pro	Glu	Glu	Lys	Pro	Asp	Ala	Thr	Pro
				470					475					480
Gln	Asn	Pro	Phe	Leu	Lys	Ser	Arg	Pro	Gln	Val	Arg	Pro	Lys	Pro
				485					490					495
Ala	Pro	Ser	Pro	Lys	Thr	Glu	Pro	Pro	Gln	Gly	Glu	Asp	Gln	Val
				500					505					510
Asp	Ile	Cys	Asn	Leu	Arg	Ser	Lys	Leu	Arg	Pro	Ala	Lys	Ser	Gln
				515					520					525
Asp	Lys	Ser	Leu	Leu	Asp	Gly	Glu	Gly	Pro	Gln	Ala	Val	Gly	Gly
				530					535					540
Gln	Asp	Val	Ala	Phe	Ser	Arg	Ser	Phe	Leu	Pro	Gly	Glu	Gly	Pro
				545					550					555
Gly	Arg	Ala	Gln	Asp	Arg	Thr	Gly	Lys	Gln	Asp	Gly	Leu	Ser	Pro
				560					565					570
Lys	Glu	Ile	Ser	Cys	Arg	Ala	Pro	Pro	Arg	Pro	Ala	Lys	Thr	Thr
				575					580					585
Asp	Pro	Val	Ser	Lys	Ser	Val	Pro	Val	Pro	Leu	Gln	Glu	Ala	Pro
				590					595					600
Gln	Gln	Arg	Pro	Val	Val	Pro	Pro	Arg	Arg	Pro	Pro	Pro	Pro	Lys
				605					610					615
Lys	Thr	Ser	Ser	Ser	Ser	Arg	Pro	Leu	Pro	Glu	Val	Arg	Gly	Pro
				620					625					630
Gln	Cys	Glu	Gly	His	Glu	Ser	Arg	Ala	Ala	Pro	Thr	Pro	Gly	Arg
				635					640					645
Ala	Leu	Leu	Val	Pro	Pro	Lys	Ala	Lys	Pro	Phe	Leu	Ser	Asn	Ser
				650					655					660
Leu	Gly	Gly	Gln	Asp	Asp	Thr	Arg	Gly	Lys	Gly	Ser	Leu	Gly	Pro
				665					670					675
Trp	Gly	Thr	Gly	Lys	Ile	Gly	Glu	Asn	Arg	Glu	Lys	Ala	Ala	Ala
				680					685					690
Ala	Ser	Val	Pro	Asn	Ala	Asp	Gly	Leu	Lys	Asp	Ser	Leu	Tyr	Val
				695					700					705

Ala	Val	Ala	Asp	Phe	Glu	Gly	Asp	Lys	Asp	Thr	Ser	Ser	Phe	Gln
				710					715					720
Glu	Gly	Thr	Val	Phe	Glu	Val	Arg	Glu	Lys	Asn	Ser	Ser	Gly	Trp
				725					730					735
Trp	Phe	Cys	Gln	Val	Leu	Ser	Gly	Ala	Pro	Ser	Trp	Glu	Gly	Trp
				740					745					750
Ile	Pro	Ser	Asn	Tyr	Leu	Arg	Lys	Lys	Pro					
				755					760					

<210> 5

<211> 398

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 5977982CD1

<400> 5

Met	Pro	Val	Pro	Ala	Ser	Trp	Pro	His	Leu	Pro	Ser	Pro	Phe	Leu
1				5					10					15
Leu	Met	Thr	Leu	Leu	Leu	Gly	Gly	Leu	Thr	Gly	Val	Ala	Gly	Glu
				20					25					30
Glu	Glu	Leu	Gln	Val	Ile	Gln	Pro	Asp	Lys	Ser	Ile	Ser	Val	Ala
				35					40					45
Ala	Gly	Glu	Ser	Ala	Thr	Leu	His	Cys	Thr	Val	Thr	Ser	Leu	Ile
				50					55					60
Pro	Val	Gly	Pro	Ile	Gln	Trp	Phe	Arg	Gly	Ala	Gly	Pro	Gly	Arg
				65					70					75
Glu	Leu	Ile	Tyr	Asn	Gln	Lys	Glu	Gly	His	Phe	Pro	Arg	Val	Thr
				80					85					90
Thr	Val	Ser	Asp	Leu	Thr	Lys	Arg	Asn	Asn	Met	Asp	Phe	Ser	Ile
				95					100					105
Arg	Ile	Ser	Asn	Ile	Thr	Pro	Ala	Asp	Ala	Gly	Thr	Tyr	Tyr	Cys
				110					115					120
Val	Lys	Phe	Arg	Lys	Gly	Ser	Pro	Asp	His	Val	Glu	Phe	Lys	Ser
				125					130					135
Gly	Ala	Gly	Thr	Glu	Leu	Ser	Val	Arg	Ala	Lys	Pro	Ser	Ala	Pro
				140					145					150
Val	Val	Ser	Gly	Pro	Ala	Ala	Arg	Ala	Thr	Pro	Gln	His	Thr	Val
				155					160					165
Ser	Phe	Thr	Cys	Glu	Ser	His	Gly	Phe	Ser	Pro	Arg	Asp	Ile	Thr
				170					175					180
Leu	Lys	Trp	Phe	Lys	Asn	Gly	Asn	Glu	Leu	Ser	Asp	Phe	Gln	Thr
				185					190					195
Asn	Val	Asp	Pro	Ala	Gly	Asp	Ser	Val	Ser	Tyr	Ser	Ile	His	Ser
				200					205					210
Thr	Ala	Lys	Val	Val	Leu	Thr	Arg	Glu	Asp	Val	His	Ser	Gln	Val
				215					220					225
Ile	Cys	Glu	Val	Ala	His	Val	Thr	Leu	Gln	Gly	Asp	Pro	Leu	Arg
				230					235					240
Gly	Thr	Ala	Asn	Leu	Ser	Glu	Thr	Ile	Arg	Val	Pro	Pro	Thr	Leu
				245					250					255
Glu	Val	Thr	Gln	Gln	Pro	Val	Arg	Ala	Glu	Asn	Gln	Val	Asn	Val
				260					265					270
Thr	Cys	Gln	Val	Arg	Lys	Phe	Tyr	Pro	Gln	Arg	Leu	Gln	Leu	Thr
				275					280					285
Trp	Leu	Glu	Asn	Gly	Asn	Val	Ser	Arg	Thr	Glu	Thr	Ala	Ser	Thr
				290					295					300
Leu	Thr	Glu	Asn	Lys	Asp	Gly	Thr	Tyr	Asn	Trp	Met	Ser	Trp	Leu
				305					310					315
Leu	Val	Asn	Val	Ser	Ala	His	Arg	Asp	Asp	Val	Lys	Leu	Thr	Cys
				320					325					330

Gln	Val	Glu	His	Asp	Gly	Gln	Pro	Ala	Val	Ser	Lys	Ser	His	Asp
				335					340					345
Leu	Lys	Val	Ser	Ala	His	Pro	Lys	Glu	Gln	Gly	Ser	Asn	Thr	Ala
				350					355					360
Pro	Gly	Pro	Ala	Leu	Ala	Ser	Ala	Ala	Pro	Leu	Leu	Ile	Ala	Phe
				365					370					375
Leu	Leu	Gly	Pro	Lys	Val	Leu	Leu	Val	Val	Gly	Val	Ser	Val	Ile
				380					385					390
Tyr	Val	Tyr	Trp	Lys	Gln	Lys	Ala							
				395										

<210> 6

<211> 416

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 6880271CD1

<400> 6

Met	Ala	Thr	Glu	Pro	Pro	Ser	Pro	Leu	Arg	Val	Glu	Ala	Pro	Gly
1				5					10					15
Pro	Pro	Glu	Met	Arg	Thr	Ser	Pro	Ala	Ile	Glu	Ser	Thr	Pro	Glu
				20					25					30
Gly	Thr	Pro	Gln	Pro	Ala	Gly	Gly	Arg	Leu	Arg	Phe	Leu	Asn	Gly
				35					40					45
Cys	Val	Pro	Leu	Ser	His	Gln	Val	Ala	Gly	His	Met	Tyr	Gly	Lys
				50					55					60
Asp	Lys	Val	Gly	Ile	Leu	Gln	His	Pro	Asp	Gly	Thr	Val	Leu	Lys
				65					70					75
Gln	Leu	Gln	Pro	Pro	Pro	Arg	Gly	Pro	Arg	Glu	Leu	Glu	Phe	Tyr
				80					85					90
Asn	Met	Val	Tyr	Ala	Ala	Asp	Cys	Phe	Asp	Gly	Val	Leu	Leu	Glu
				95					100					105
Leu	Arg	Lys	Tyr	Leu	Pro	Lys	Tyr	Tyr	Gly	Ile	Trp	Ser	Pro	Pro
				110					115					120
Thr	Ala	Pro	Asn	Asp	Leu	Tyr	Leu	Lys	Leu	Glu	Asp	Val	Thr	His
				125					130					135
Lys	Phe	Asn	Lys	Pro	Cys	Ile	Met	Asp	Val	Lys	Ile	Gly	Gln	Lys
				140					145					150
Ser	Tyr	Asp	Pro	Phe	Ala	Ser	Ser	Glu	Lys	Ile	Gln	Gln	Gln	Val
				155					160					165
Ser	Lys	Tyr	Pro	Leu	Met	Glu	Glu	Ile	Gly	Phe	Leu	Val	Leu	Gly
				170					175					180
Met	Arg	Val	Tyr	His	Val	His	Ser	Asp	Ser	Tyr	Glu	Thr	Glu	Asn
				185					190					195
Gln	His	Tyr	Gly	Arg	Ser	Leu	Thr	Lys	Glu	Thr	Ile	Lys	Asp	Gly
				200					205					210
Val	Ser	Arg	Phe	Phe	His	Asn	Gly	Tyr	Cys	Leu	Arg	Lys	Asp	Ala
				215					220					225
Val	Ala	Ala	Ser	Ile	Gln	Lys	Ile	Glu	Lys	Ile	Leu	Gln	Trp	Phe
				230					235					240
Glu	Asn	Gln	Lys	Gln	Leu	Asn	Phe	Tyr	Ala	Ser	Ser	Leu	Leu	Phe
				245					250					255
Val	Tyr	Glu	Gly	Ser	Ser	Gln	Pro	Thr	Thr	Thr	Lys	Leu	Asn	Asp
				260					265					270
Arg	Thr	Leu	Ala	Glu	Lys	Phe	Leu	Ser	Lys	Gly	Gln	Leu	Ser	Asp
				275					280					285
Thr	Glu	Val	Leu	Glu	Tyr	Asn	Asn	Asn	Phe	His	Val	Leu	Ser	Ser
				290					295					300
Thr	Ala	Asn	Gly	Lys	Ile	Glu	Ser	Ser	Val	Gly	Lys	Ser	Leu	Ser
				305					310					315

Lys	Met	Tyr	Ala	Arg	His	Arg	Lys	Ile	Tyr	Thr	Lys	Lys	His	His
				320					325					330
Ser	Gln	Thr	Ser	Leu	Lys	Val	Glu	Asn	Leu	Glu	Gln	Asp	Asn	Gly
				335					340					345
Trp	Lys	Ser	Met	Ser	Gln	Glu	His	Leu	Asn	Gly	Asn	Val	Leu	Ser
				350					355					360
Gln	Leu	Glu	Lys	Val	Phe	Tyr	His	Leu	Pro	Thr	Gly	Cys	Gln	Glu
				365					370					375
Ile	Ala	Glu	Val	Glu	Val	Arg	Met	Ile	Asp	Phe	Ala	His	Val	Phe
				380					385					390
Pro	Ser	Asn	Thr	Ile	Asp	Glu	Gly	Tyr	Val	Tyr	Gly	Leu	Lys	His
				395					400					405
Leu	Ile	Ser	Val	Leu	Arg	Ser	Ile	Leu	Asp	Asn				
				410					415					

<210> 7

<211> 359

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 2378756CD1

<400> 7

Met	Ala	Gly	Tyr	Lys	Pro	Val	Ala	Ile	Gln	Thr	Tyr	Pro	Ile	Leu
1				5					10					15
Gly	Glu	Lys	Ile	Thr	Gln	Asp	Thr	Leu	Tyr	Trp	Asn	Asn	Tyr	Lys
				20					25					30
Thr	Pro	Val	Gln	Ile	Lys	Glu	Phe	Gly	Ala	Val	Ser	Lys	Val	Asp
				35					40					45
Phe	Ser	Pro	Gln	Pro	Pro	Tyr	Asn	Tyr	Ala	Val	Thr	Ala	Ser	Ser
				50					55					60
Arg	Ile	His	Ile	Tyr	Gly	Arg	Tyr	Ser	Gln	Glu	Pro	Ile	Lys	Thr
				65					70					75
Phe	Ser	Arg	Phe	Lys	Asp	Thr	Ala	Tyr	Cys	Ala	Thr	Phe	Arg	Gln
				80					85					90
Asp	Gly	Arg	Leu	Leu	Val	Ala	Gly	Ser	Glu	Asp	Gly	Gly	Val	Gln
				95					100					105
Leu	Phe	Asp	Ile	Ser	Gly	Arg	Ala	Pro	Leu	Arg	Gln	Phe	Glu	Gly
				110					115					120
His	Thr	Lys	Ala	Val	His	Thr	Val	Asp	Phe	Thr	Ala	Asp	Lys	Tyr
				125					130					135
His	Val	Val	Ser	Gly	Ala	Asp	Asp	Tyr	Thr	Val	Lys	Leu	Trp	Asp
				140					145					150
Ile	Pro	Asn	Ser	Lys	Glu	Ile	Leu	Thr	Phe	Lys	Glu	His	Ser	Asp
				155					160					165
Tyr	Val	Arg	Cys	Gly	Cys	Ala	Ser	Lys	Leu	Asn	Pro	Asp	Leu	Phe
				170					175					180
Ile	Thr	Gly	Ser	Tyr	Asp	His	Thr	Val	Lys	Met	Phe	Asp	Ala	Arg
				185					190					195
Thr	Ser	Glu	Ser	Val	Leu	Ser	Val	Glu	His	Gly	Gln	Pro	Val	Glu
				200					205					210
Ser	Val	Leu	Leu	Phe	Pro	Ser	Gly	Gly	Leu	Leu	Val	Ser	Ala	Gly
				215					220					225
Gly	Arg	Tyr	Val	Lys	Val	Trp	Asp	Met	Leu	Lys	Gly	Gly	Gln	Leu
				230					235					240
Leu	Val	Ser	Leu	Lys	Asn	His	His	Lys	Thr	Val	Thr	Cys	Leu	Cys
				245					250					255
Leu	Ser	Ser	Ser	Gly	Gln	Trp	Leu	Leu	Ser	Gly	Ser	Leu	Asp	Arg
				260					265					270
Lys	Val	Lys	Val	Tyr	Ser	Thr	Thr	Ser	Tyr	Lys	Val	Val	His	Ser
				275					280					285

Phe	Asp	Tyr	Ala	Ala	Ser	Ile	Leu	Ser	Leu	Ala	Leu	Ala	His	Glu
				290					295					300
Asp	Glu	Thr	Ile	Val	Val	Gly	Met	Thr	Asn	Gly	Ile	Leu	Ser	Val
				305					310					315
Lys	His	Arg	Lys	Ser	Glu	Ala	Lys	Lys	Glu	Ser	Leu	Pro	Arg	Arg
				320					325					330
Arg	Arg	Pro	Ala	Tyr	Arg	Thr	Phe	Ile	Lys	Gly	Lys	Asn	Tyr	Met
				335					340					345
Lys	Gln	Arg	Val	Phe	Val	His	Phe	Ser	Tyr	Leu	Phe	Lys	Gly	
				350					355					

<210> 8

<211> 1243

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 1861527CD1

<400> 8

Met	Leu	Ala	Ser	Pro	Ala	Thr	Glu	Thr	Thr	Val	Leu	Met	Ser	Gln
1				5					10					15
Thr	Glu	Ala	Asp	Leu	Ala	Leu	Arg	Pro	Pro	Pro	Pro	Leu	Gly	Thr
				20					25					30
Ala	Gly	Gln	Pro	Arg	Leu	Gly	Pro	Pro	Pro	Arg	Arg	Ala	Arg	Arg
				35					40					45
Phe	Ser	Gly	Lys	Ala	Glu	Pro	Arg	Pro	Arg	Ser	Ser	Arg	Leu	Ser
				50					55					60
Arg	Arg	Ser	Ser	Val	Asp	Leu	Gly	Leu	Leu	Ser	Ser	Trp	Ser	Leu
				65					70					75
Pro	Ala	Ser	Pro	Ala	Pro	Asp	Pro	Pro	Asp	Pro	Pro	Asp	Ser	Ala
				80					85					90
Gly	Pro	Gly	Pro	Ala	Arg	Ser	Pro	Pro	Pro	Ser	Ser	Lys	Glu	Pro
				95					100					105
Pro	Glu	Gly	Thr	Trp	Thr	Glu	Gly	Ala	Pro	Val	Lys	Ala	Ala	Glu
				110					115					120
Asp	Ser	Ala	Arg	Pro	Glu	Leu	Pro	Asp	Ser	Ala	Val	Gly	Pro	Gly
				125					130					135
Ser	Arg	Glu	Pro	Leu	Arg	Val	Pro	Glu	Ala	Val	Ala	Leu	Glu	Arg
				140					145					150
Arg	Arg	Glu	Gln	Glu	Glu	Lys	Glu	Asp	Met	Glu	Thr	Gln	Ala	Val
				155					160					165
Ala	Thr	Ser	Pro	Asp	Gly	Arg	Tyr	Leu	Lys	Phe	Asp	Ile	Glu	Ile
				170					175					180
Gly	Arg	Gly	Ser	Phe	Lys	Thr	Val	Tyr	Arg	Gly	Leu	Asp	Thr	Asp
				185					190					195
Thr	Thr	Val	Glu	Val	Ala	Trp	Cys	Glu	Leu	Gln	Thr	Arg	Lys	Leu
				200					205					210
Ser	Arg	Ala	Glu	Arg	Gln	Arg	Phe	Ser	Glu	Glu	Val	Glu	Met	Leu
				215					220					225
Lys	Gly	Leu	Gln	His	Pro	Asn	Ile	Val	Arg	Phe	Tyr	Asp	Ser	Trp
				230					235					240
Lys	Ser	Val	Leu	Arg	Gly	Gln	Val	Cys	Ile	Val	Leu	Val	Thr	Glu
				245					250					255
Leu	Met	Thr	Ser	Gly	Thr	Leu	Lys	Thr	Tyr	Leu	Arg	Arg	Phe	Arg
				260					265					270
Glu	Met	Lys	Pro	Arg	Val	Leu	Gln	Arg	Trp	Ser	Arg	Gln	Ile	Leu
				275					280					285
Arg	Gly	Leu	His	Phe	Leu	His	Ser	Arg	Val	Pro	Pro	Ile	Leu	His
				290					295					300
Arg	Asp	Leu	Lys	Cys	Asp	Asn	Val	Phe	Ile	Thr	Gly	Pro	Ser	Gly
				305					310					315

Ser	Val	Lys	Ile	Gly	Asp	Leu	Gly	Leu	Ala	Thr	Leu	Lys	Arg	Ala
				320					325					330
Ser	Phe	Ala	Lys	Ser	Val	Ile	Gly	Thr	Pro	Glu	Phe	Met	Ala	Pro
				335					340					345
Glu	Met	Tyr	Glu	Glu	Lys	Tyr	Asp	Glu	Ala	Val	Asp	Val	Tyr	Ala
				350					355					360
Phe	Gly	Met	Cys	Met	Leu	Glu	Met	Ala	Thr	Ser	Glu	Tyr	Pro	Tyr
				365					370					375
Ser	Glu	Cys	Gln	Asn	Ala	Ala	Gln	Ile	Tyr	Arg	Lys	Val	Thr	Ser
				380					385					390
Gly	Arg	Lys	Pro	Asn	Ser	Phe	His	Lys	Val	Lys	Ile	Pro	Glu	Val
				395					400					405
Lys	Glu	Ile	Ile	Glu	Gly	Cys	Ile	Arg	Thr	Asp	Lys	Asn	Glu	Arg
				410					415					420
Phe	Thr	Ile	Gln	Asp	Leu	Leu	Ala	His	Ala	Phe	Phe	Arg	Glu	Glu
				425					430					435
Arg	Gly	Val	His	Val	Glu	Leu	Ala	Glu	Glu	Asp	Asp	Gly	Glu	Lys
				440					445					450
Pro	Gly	Leu	Lys	Leu	Trp	Leu	Arg	Met	Glu	Asp	Ala	Arg	Arg	Gly
				455					460					465
Gly	Arg	Pro	Arg	Asp	Asn	Gln	Ala	Ile	Glu	Phe	Leu	Phe	Gln	Leu
				470					475					480
Gly	Arg	Asp	Ala	Ala	Glu	Glu	Val	Ala	Gln	Glu	Met	Val	Ala	Leu
				485					490					495
Gly	Leu	Val	Cys	Glu	Ala	Asp	Tyr	Gln	Pro	Val	Ala	Arg	Ala	Val
				500					505					510
Arg	Glu	Arg	Val	Ala	Ala	Ile	Gln	Arg	Lys	Arg	Glu	Lys	Leu	Arg
				515					520					525
Lys	Ala	Arg	Glu	Leu	Glu	Ala	Leu	Pro	Pro	Glu	Pro	Gly	Pro	Pro
				530					535					540
Pro	Ala	Thr	Val	Pro	Met	Ala	Pro	Gly	Pro	Pro	Ser	Val	Phe	Pro
				545					550					555
Pro	Glu	Pro	Glu	Glu	Pro	Glu	Ala	Asp	Gln	His	Gln	Pro	Phe	Leu
				560					565					570
Phe	Arg	His	Ala	Ser	Tyr	Ser	Ser	Thr	Thr	Ser	Asp	Cys	Glu	Thr
				575					580					585
Asp	Gly	Tyr	Leu	Ser	Ser	Ser	Gly	Phe	Leu	Asp	Ala	Ser	Asp	Pro
				590					595					600
Ala	Leu	Gln	Pro	Pro	Gly	Gly	Val	Pro	Ser	Ser	Leu	Ala	Glu	Ser
				605					610					615
His	Leu	Cys	Leu	Pro	Ser	Ala	Phe	Ala	Leu	Ser	Ile	Pro	Arg	Ser
				620					625					630
Gly	Pro	Gly	Ser	Asp	Phe	Ser	Pro	Gly	Asp	Ser	Tyr	Ala	Ser	Asp
				635					640					645
Ala	Ala	Ser	Gly	Leu	Ser	Asp	Val	Gly	Glu	Gly	Met	Gly	Gln	Met
				650					655					660
Arg	Arg	Pro	Pro	Gly	Arg	Asn	Leu	Arg	Arg	Arg	Pro	Arg	Ser	Arg
				665					670					675
Leu	Arg	Val	Thr	Ser	Val	Ser	Asp	Gln	Asn	Asp	Arg	Val	Val	Glu
				680					685					690
Cys	Gln	Leu	Gln	Thr	His	Asn	Ser	Lys	Met	Val	Thr	Phe	Arg	Phe
				695					700					705
Asp	Leu	Asp	Gly	Asp	Ser	Pro	Glu	Glu	Ile	Ala	Ala	Ala	Met	Val
				710					715					720
Tyr	Asn	Glu	Phe	Ile	Leu	Pro	Ser	Glu	Arg	Asp	Gly	Phe	Leu	Arg
				725					730					735
Arg	Ile	Arg	Glu	Ile	Ile	Gln	Arg	Val	Glu	Thr	Leu	Leu	Lys	Arg
				740					745					750
Asp	Thr	Gly	Pro	Met	Glu	Ala	Ala	Glu	Asp	Thr	Leu	Ser	Pro	Gln
				755					760					765
Glu	Glu	Pro	Ala	Pro	Leu	Pro	Ala	Leu	Pro	Val	Pro	Leu	Pro	Asp
				770					775					780
Pro	Ser	Asn	Glu	Glu	Leu	Gln	Ser	Ser	Thr	Ser	Leu	Glu	His	Arg

				785					790					795
Ser	Trp	Thr	Ala	Phe	Ser	Thr	Ser	Ser	Ser	Ser	Pro	Gly	Thr	Pro
				800					805					810
Leu	Ser	Pro	Gly	Asn	Pro	Phe	Ser	Pro	Gly	Thr	Pro	Ile	Ser	Pro
				815					820					825
Gly	Pro	Ile	Phe	Pro	Ile	Thr	Ser	Pro	Pro	Cys	His	Pro	Ser	Pro
				830					835					840
Ser	Pro	Phe	Ser	Pro	Ile	Ser	Ser	Gln	Val	Ser	Ser	Asn	Pro	Ser
				845					850					855
Pro	His	Pro	Thr	Ser	Ser	Pro	Leu	Pro	Phe	Ser	Ser	Ser	Thr	Pro
				860					865					870
Glu	Phe	Pro	Val	Pro	Leu	Ser	Gln	Cys	Pro	Trp	Ser	Ser	Leu	Pro
				875					880					885
Thr	Thr	Ser	Pro	Pro	Thr	Phe	Ser	Pro	Thr	Cys	Ser	Gln	Val	Thr
				890					895					900
Leu	Ser	Ser	Pro	Phe	Phe	Pro	Pro	Cys	Pro	Ser	Thr	Ser	Ser	Phe
				905					910					915
Pro	Ser	Thr	Thr	Ala	Ala	Pro	Leu	Leu	Ser	Leu	Ala	Ser	Ala	Phe
				920					925					930
Ser	Leu	Ala	Val	Met	Thr	Val	Ala	Gln	Ser	Leu	Leu	Ser	Pro	Ser
				935					940					945
Pro	Gly	Leu	Leu	Ser	Gln	Ser	Pro	Pro	Ala	Pro	Pro	Ser	Pro	Leu
				950					955					960
Pro	Ser	Leu	Pro	Leu	Pro	Pro	Pro	Val	Ala	Pro	Gly	Gly	Gln	Glu
				965					970					975
Ser	Pro	Ser	Pro	His	Thr	Ala	Glu	Val	Glu	Ser	Glu	Ala	Ser	Pro
				980					985					990
Pro	Pro	Ala	Arg	Pro	Leu	Pro	Gly	Glu	Ala	Arg	Leu	Ala	Pro	Ile
				995					1000					1005
Ser	Glu	Glu	Gly	Lys	Pro	Gln	Leu	Val	Gly	Arg	Phe	Gln	Val	Thr
				1010					1015					1020
Ser	Ser	Lys	Glu	Pro	Ala	Glu	Pro	Leu	Pro	Leu	Gln	Pro	Thr	Ser
				1025					1030					1035
Pro	Thr	Leu	Ser	Gly	Ser	Pro	Lys	Pro	Ser	Thr	Pro	Gln	Leu	Thr
				1040					1045					1050
Ser	Glu	Ser	Ser	Asp	Thr	Glu	Asp	Ser	Ala	Gly	Gly	Gly	Pro	Glu
				1055					1060					1065
Thr	Arg	Glu	Ala	Leu	Ala	Glu	Ser	Asp	Arg	Ala	Ala	Glu	Gly	Leu
				1070					1075					1080
Gly	Ala	Gly	Val	Glu	Glu	Glu	Gly	Asp	Asp	Gly	Lys	Glu	Pro	Gln
				1085					1090					1095
Val	Gly	Gly	Ser	Pro	Gln	Pro	Leu	Ser	His	Pro	Ser	Pro	Val	Trp
				1100					1105					1110
Met	Asn	Tyr	Ser	Tyr	Ser	Ser	Leu	Cys	Leu	Ser	Ser	Glu	Glu	Ser
				1115					1120					1125
Glu	Ser	Ser	Gly	Glu	Asp	Glu	Glu	Phe	Trp	Ala	Glu	Leu	Gln	Ser
				1130					1135					1140
Leu	Arg	Gln	Lys	His	Leu	Ser	Glu	Val	Glu	Thr	Leu	Gln	Thr	Leu
				1145					1150					1155
Gln	Lys	Lys	Glu	Ile	Glu	Asp	Leu	Tyr	Ser	Arg	Leu	Gly	Lys	Gln
				1160					1165					1170
Pro	Pro	Pro	Gly	Ile	Val	Ala	Pro	Ala	Ala	Met	Leu	Ser	Ser	Arg
				1175					1180					1185
Gln	Arg	Arg	Leu	Ser	Lys	Gly	Ser	Phe	Pro	Thr	Ser	Arg	Arg	Asn
				1190					1195					1200
Ser	Leu	Gln	Arg	Ser	Glu	Pro	Pro	Gly	Pro	Gly	Ile	Met	Arg	Arg
				1205					1210					1215
Asn	Ser	Leu	Ser	Gly	Ser	Ser	Thr	Gly	Ser	Gln	Glu	Gln	Arg	Ala
				1220					1225					1230
Ser	Lys	Gly	Val	Thr	Phe	Ala	Gly	Asp	Val	Gly	Arg	Met		
				1235					1240					

<210> 9

<211> 373

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 2921356CD1

<400> 9

Met	Asp	Ser	Gly	Arg	Lys	Asn	Arg	Pro	Pro	Phe	Pro	Trp	Phe	Gly
1				5					10					15
Met	Asp	Ile	Gly	Gly	Thr	Leu	Val	Lys	Leu	Val	Tyr	Phe	Glu	Pro
				20					25					30
Lys	Asp	Ile	Thr	Ala	Glu	Glu	Glu	Gln	Glu	Glu	Val	Glu	Asn	Leu
				35					40					45
Lys	Ser	Ile	Arg	Lys	Tyr	Leu	Thr	Ser	Asn	Thr	Ala	Tyr	Gly	Lys
				50					55					60
Thr	Gly	Ile	Arg	Asp	Val	His	Leu	Glu	Leu	Lys	Asn	Leu	Thr	Met
				65					70					75
Cys	Gly	Arg	Lys	Gly	Asn	Leu	His	Phe	Ile	Arg	Phe	Pro	Ser	Cys
				80					85					90
Ala	Met	His	Arg	Phe	Ile	Gln	Met	Gly	Ser	Glu	Lys	Asn	Phe	Ser
				95					100					105
Ser	Leu	His	Thr	Thr	Leu	Cys	Ala	Thr	Gly	Gly	Gly	Ala	Phe	Lys
				110					115					120
Phe	Glu	Glu	Asp	Phe	Arg	Met	Ile	Ala	Asp	Leu	Gln	Leu	His	Lys
				125					130					135
Leu	Asp	Glu	Leu	Asp	Cys	Leu	Ile	Gln	Gly	Leu	Leu	Tyr	Val	Asp
				140					145					150
Ser	Val	Gly	Phe	Asn	Gly	Lys	Pro	Glu	Cys	Tyr	Tyr	Phe	Glu	Asn
				155					160					165
Pro	Thr	Asn	Pro	Glu	Leu	Cys	Gln	Lys	Lys	Pro	Tyr	Cys	Leu	Asp
				170					175					180
Asn	Pro	Tyr	Pro	Met	Leu	Leu	Val	Asn	Met	Gly	Ser	Gly	Val	Ser
				185					190					195
Ile	Leu	Ala	Val	Tyr	Ser	Lys	Asp	Asn	Tyr	Lys	Arg	Val	Thr	Gly
				200					205					210
Thr	Ser	Leu	Gly	Gly	Gly	Thr	Phe	Leu	Gly	Leu	Cys	Cys	Leu	Leu
				215					220					225
Thr	Gly	Cys	Glu	Thr	Phe	Glu	Glu	Ala	Leu	Glu	Met	Ala	Ala	Lys
				230					235					240
Gly	Asp	Ser	Thr	Asn	Val	Asp	Lys	Leu	Val	Lys	Asp	Ile	Tyr	Gly
				245					250					255
Gly	Asp	Tyr	Glu	Arg	Phe	Gly	Leu	Gln	Gly	Ser	Ala	Val	Ala	Ser
				260					265					270
Ser	Phe	Gly	Asn	Met	Met	Ser	Lys	Glu	Lys	Arg	Asp	Ser	Ile	Ser
				275					280					285
Lys	Glu	Asp	Leu	Ala	Arg	Ala	Thr	Leu	Val	Thr	Ile	Thr	Asn	Asn
				290					295					300
Ile	Gly	Ser	Ile	Ala	Arg	Met	Cys	Ala	Leu	Asn	Glu	Asn	Ile	Asp
				305					310					315
Arg	Val	Val	Phe	Val	Gly	Asn	Phe	Leu	Arg	Ile	Asn	Met	Val	Ser
				320					325					330
Met	Lys	Leu	Leu	Ala	Tyr	Ala	Met	Asp	Phe	Trp	Ser	Lys	Gly	Gln
				335					340					345
Leu	Lys	Ala	Leu	Phe	Leu	Glu	His	Glu	Gly	Tyr	Phe	Gly	Ala	Val
				350					355					360
Gly	Ala	Leu	Leu	Glu	Leu	Phe	Lys	Met	Thr	Asp	Asp	Lys		
				365					370					

<210> 10

<211> 548

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7386170CD1

<400> 10

Met	Pro	Ser	Arg	Thr	Gly	Pro	Lys	Met	Glu	Gly	Ser	Gly	Gly	Arg
1				5					10					15
Val	Arg	Leu	Lys	Ala	His	Tyr	Gly	Gly	Asp	Ile	Phe	Ile	Thr	Ser
				20					25					30
Val	Asp	Ala	Ala	Thr	Thr	Phe	Glu	Glu	Leu	Cys	Glu	Glu	Val	Arg
				35					40					45
Asp	Met	Cys	Arg	Leu	His	Gln	Gln	His	Pro	Leu	Thr	Leu	Lys	Trp
				50					55					60
Val	Asp	Ser	Glu	Gly	Asp	Pro	Cys	Thr	Val	Ser	Ser	Gln	Met	Glu
				65					70					75
Leu	Glu	Glu	Ala	Phe	Arg	Leu	Ala	Arg	Gln	Cys	Arg	Asp	Glu	Gly
				80					85					90
Leu	Ile	Ile	His	Val	Phe	Pro	Ser	Thr	Pro	Glu	Gln	Pro	Gly	Leu
				95					100					105
Pro	Cys	Pro	Gly	Glu	Asp	Lys	Ser	Ile	Tyr	Arg	Arg	Gly	Ala	Arg
				110					115					120
Arg	Trp	Arg	Lys	Leu	Tyr	Arg	Ala	Asn	Gly	His	Leu	Phe	Gln	Ala
				125					130					135
Lys	Arg	Phe	Asn	Arg	Asp	Ser	Val	Met	Pro	Ser	Gln	Glu	Pro	Pro
				140					145					150
Val	Asp	Asp	Lys	Asn	Glu	Asp	Ala	Asp	Leu	Pro	Ser	Glu	Glu	Thr
				155					160					165
Asp	Gly	Ile	Ala	Tyr	Ile	Ser	Ser	Ser	Arg	Lys	His	Asp	Ser	Ile
				170					175					180
Lys	Asp	Asp	Ser	Glu	Asp	Leu	Lys	Pro	Val	Ile	Asp	Gly	Met	Asp
				185					190					195
Gly	Ile	Lys	Ile	Ser	Gln	Gly	Leu	Gly	Leu	Gln	Asp	Phe	Asp	Leu
				200					205					210
Ile	Arg	Val	Ile	Gly	Arg	Gly	Ser	Tyr	Ala	Lys	Val	Leu	Leu	Val
				215					220					225
Arg	Leu	Lys	Lys	Asn	Asp	Gln	Ile	Tyr	Ala	Met	Lys	Val	Val	Lys
				230					235					240
Lys	Glu	Leu	Val	His	Asp	Asp	Glu	Asp	Ile	Asp	Trp	Val	Gln	Thr
				245					250					255
Glu	Lys	His	Val	Phe	Glu	Gln	Ala	Ser	Ser	Asn	Pro	Phe	Leu	Val
				260					265					270
Gly	Leu	His	Ser	Cys	Phe	Gln	Thr	Thr	Ser	Arg	Leu	Phe	Leu	Val
				275					280					285
Ile	Glu	Tyr	Val	Asn	Gly	Gly	Asp	Leu	Met	Phe	His	Met	Gln	Arg
				290					295					300
Gln	Arg	Lys	Leu	Pro	Glu	Glu	His	Ala	Arg	Phe	Tyr	Ala	Ala	Glu
				305					310					315
Ile	Cys	Ile	Ala	Leu	Asn	Phe	Leu	His	Glu	Arg	Gly	Ile	Ile	Tyr
				320					325					330
Arg	Asp	Leu	Lys	Leu	Asp	Asn	Val	Leu	Leu	Asp	Ala	Asp	Gly	His
				335					340					345
Ile	Lys	Leu	Thr	Asp	Tyr	Gly	Met	Cys	Lys	Glu	Gly	Leu	Gly	Pro
				350					355					360
Gly	Asp	Thr	Thr	Ser	Thr	Phe	Cys	Gly	Thr	Pro	Asn	Tyr	Ile	Ala
				365					370					375
Pro	Glu	Ile	Leu	Arg	Gly	Glu	Glu	Tyr	Gly	Phe	Ser	Val	Asp	Trp
				380					385					390
Trp	Ala	Leu	Gly	Val	Leu	Met	Phe	Glu	Met	Met	Ala	Gly	Arg	Ser
				395					400					405
Pro	Phe	Asp	Ile	Ile	Thr	Asp	Asn	Pro	Asp	Met	Asn	Thr	Glu	Asp
				410					415					420

Tyr	Leu	Phe	Gln	Val	Ile	Leu	Glu	Lys	Pro	Ile	Arg	Ile	Pro	Arg
				425					430					435
Phe	Leu	Ser	Val	Lys	Ala	Ser	His	Val	Leu	Lys	Gly	Phe	Leu	Asn
				440					445					450
Lys	Asp	Pro	Lys	Glu	Arg	Leu	Gly	Cys	Arg	Pro	Gln	Thr	Gly	Phe
				455					460					465
Ser	Asp	Ile	Lys	Ser	His	Ala	Phe	Phe	Arg	Ser	Ile	Asp	Trp	Asp
				470					475					480
Leu	Leu	Glu	Lys	Lys	Gln	Ala	Leu	Pro	Pro	Phe	Gln	Pro	Gln	Ile
				485					490					495
Thr	Asp	Asp	Tyr	Gly	Leu	Asp	Asn	Phe	Asp	Thr	Gln	Phe	Thr	Ser
				500					505					510
Glu	Pro	Val	Gln	Leu	Thr	Pro	Asp	Asp	Glu	Asp	Ala	Ile	Lys	Arg
				515					520					525
Ile	Asp	Gln	Ser	Glu	Phe	Glu	Gly	Phe	Glu	Tyr	Ile	Asn	Pro	Leu
				530					535					540
Leu	Leu	Ser	Thr	Glu	Glu	Ser	Val							
				545										

<210> 11

<211> 1093

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7481206CD1

<400> 11

Met	Ala	Gly	Ala	Ala	Gly	Leu	Thr	Ala	Glu	Val	Ser	Trp	Lys	Val
1				5					10					15
Leu	Glu	Arg	Arg	Ala	Arg	Thr	Lys	Arg	Ser	Gly	Ser	Val	Tyr	Glu
				20					25					30
Pro	Leu	Lys	Ser	Ile	Asn	Leu	Pro	Arg	Pro	Asp	Asn	Glu	Thr	Leu
				35					40					45
Trp	Asp	Lys	Leu	Asp	His	Tyr	Tyr	Arg	Ile	Val	Lys	Ser	Thr	Leu
				50					55					60
Leu	Leu	Tyr	Gln	Ser	Pro	Thr	Thr	Gly	Leu	Phe	Pro	Thr	Lys	Thr
				65					70					75
Cys	Gly	Gly	Asp	Gln	Lys	Ala	Lys	Ile	Gln	Asp	Ser	Leu	Tyr	Cys
				80					85					90
Ala	Ala	Gly	Ala	Trp	Ala	Leu	Ala	Leu	Ala	Tyr	Arg	Arg	Ile	Asp
				95					100					105
Asp	Asp	Lys	Gly	Arg	Thr	His	Glu	Leu	Glu	His	Ser	Ala	Ile	Lys
				110					115					120
Cys	Met	Arg	Gly	Ile	Leu	Tyr	Cys	Tyr	Met	Arg	Gln	Ala	Asp	Lys
				125					130					135
Val	Gln	Gln	Phe	Lys	Gln	Asp	Pro	Arg	Pro	Thr	Thr	Cys	Leu	His
				140					145					150
Ser	Val	Phe	Asn	Val	His	Thr	Gly	Asp	Glu	Leu	Leu	Ser	Tyr	Glu
				155					160					165
Glu	Tyr	Gly	His	Leu	Gln	Ile	Asn	Ala	Val	Ser	Leu	Tyr	Leu	Leu
				170					175					180
Tyr	Leu	Val	Glu	Met	Ile	Ser	Ser	Gly	Leu	Gln	Ile	Ile	Tyr	Asn
				185					190					195
Thr	Asp	Glu	Val	Ser	Phe	Ile	Gln	Asn	Leu	Val	Phe	Cys	Val	Glu
				200					205					210
Arg	Val	Tyr	Arg	Val	Pro	Asp	Phe	Gly	Val	Trp	Glu	Arg	Gly	Ser
				215					220					225
Lys	Tyr	Asn	Asn	Gly	Ser	Thr	Glu	Leu	His	Ser	Ser	Ser	Val	Gly
				230					235					240
Leu	Ala	Lys	Ala	Ala	Leu	Glu	Ala	Ile	Asn	Gly	Phe	Asn	Leu	Phe
				245					250					255

Gly	Asn	Gln	Gly	Cys	Ser	Trp	Ser	Val	Ile	Phe	Val	Asp	Leu	Asp
				260					265					270
Ala	His	Asn	Arg	Asn	Arg	Gln	Thr	Leu	Cys	Ser	Leu	Leu	Pro	Arg
				275					280					285
Glu	Ser	Arg	Ser	His	Asn	Thr	Asp	Ala	Ala	Leu	Leu	Pro	Cys	Ile
				290					295					300
Ser	Tyr	Pro	Ala	Phe	Ala	Leu	Asp	Asp	Glu	Val	Leu	Phe	Ser	Gln
				305					310					315
Thr	Leu	Asp	Lys	Val	Val	Arg	Lys	Leu	Lys	Gly	Lys	Tyr	Gly	Phe
				320					325					330
Lys	Arg	Phe	Leu	Arg	Asp	Gly	Tyr	Arg	Thr	Ser	Leu	Glu	Asp	Pro
				335					340					345
Asn	Arg	Cys	Tyr	Tyr	Lys	Pro	Ala	Glu	Ile	Lys	Leu	Phe	Asp	Gly
				350					355					360
Ile	Glu	Cys	Glu	Phe	Pro	Ile	Phe	Phe	Leu	Tyr	Met	Met	Ile	Asp
				365					370					375
Gly	Val	Phe	Arg	Gly	Asn	Pro	Lys	Gln	Val	Gln	Glu	Tyr	Gln	Asp
				380					385					390
Leu	Leu	Thr	Pro	Val	Leu	His	His	Thr	Thr	Glu	Gly	Tyr	Pro	Val
				395					400					405
Val	Pro	Lys	Tyr	Tyr	Tyr	Val	Pro	Ala	Asp	Phe	Val	Glu	Tyr	Glu
				410					415					420
Lys	Asn	Asn	Pro	Gly	Ser	Gln	Lys	Arg	Phe	Pro	Ser	Asn	Cys	Gly
				425					430					435
Arg	Asp	Gly	Lys	Leu	Phe	Leu	Trp	Gly	Gln	Ala	Leu	Tyr	Ile	Ile
				440					445					450
Ala	Lys	Leu	Leu	Ala	Asp	Glu	Leu	Ile	Ser	Pro	Lys	Asp	Ile	Asp
				455					460					465
Pro	Val	Gln	Arg	Tyr	Val	Pro	Leu	Lys	Asp	Gln	Arg	Asn	Val	Ser
				470					475					480
Met	Arg	Phe	Ser	Asn	Gln	Gly	Pro	Leu	Glu	Asn	Asp	Leu	Val	Val
				485					490					495
His	Val	Ala	Leu	Ile	Ala	Glu	Ser	Gln	Arg	Leu	Gln	Val	Phe	Leu
				500					505					510
Asn	Thr	Tyr	Gly	Ile	Gln	Thr	Gln	Thr	Pro	Gln	Gln	Val	Glu	Pro
				515					520					525
Ile	Gln	Ile	Trp	Pro	Gln	Gln	Glu	Leu	Val	Lys	Ala	Tyr	Leu	Gln
				530					535					540
Leu	Gly	Ile	Asn	Glu	Lys	Leu	Gly	Leu	Ser	Gly	Arg	Pro	Asp	Arg
				545					550					555
Pro	Ile	Gly	Cys	Leu	Gly	Thr	Ser	Lys	Ile	Tyr	Arg	Ile	Leu	Gly
				560					565					570
Lys	Thr	Val	Val	Cys	Tyr	Pro	Ile	Ile	Phe	Asp	Leu	Ser	Asp	Phe
				575					580					585
Tyr	Met	Ser	Gln	Asp	Val	Phe	Leu	Leu	Ile	Asp	Asp	Ile	Lys	Asn
				590					595					600
Ala	Leu	Gln	Phe	Ile	Lys	Gln	Tyr	Trp	Lys	Met	His	Gly	Arg	Pro
				605					610					615
Leu	Phe	Leu	Val	Leu	Ile	Arg	Glu	Asp	Asn	Ile	Arg	Gly	Ser	Arg
				620					625					630
Phe	Asn	Pro	Ile	Leu	Asp	Met	Leu	Ala	Ala	Leu	Lys	Lys	Gly	Ile
				635					640					645
Ile	Gly	Gly	Val	Lys	Val	His	Val	Asp	Arg	Leu	Gln	Thr	Leu	Ile
				650					655					660
Ser	Gly	Ala	Val	Val	Glu	Gln	Leu	Asp	Phe	Leu	Arg	Ile	Ser	Asp
				665					670					675
Thr	Glu	Glu	Leu	Pro	Glu	Phe	Lys	Ser	Phe	Glu	Glu	Leu	Glu	Pro
				680					685					690
Pro	Lys	His	Ser	Lys	Val	Lys	Arg	Gln	Ser	Ser	Thr	Pro	Ser	Ala
				695					700					705
Pro	Glu	Leu	Gly	Gln	Gln	Pro	Asp	Val	Asn	Ile	Ser	Glu	Trp	Lys
				710					715					720
Asp	Lys	Pro	Thr	His	Glu	Ile	Leu	Gln	Lys	Leu	Asn	Asp	Cys	Ser

Cys	Leu	Ala	Ser	Gln	Ala	Ile	Leu	Leu	Gly	Ile	Leu	Leu	Lys	Arg	725	730	735
															740	745	750
Glu	Gly	Pro	Asn	Phe	Ile	Thr	Lys	Glu	Gly	Thr	Val	Ser	Asp	His			
															755	760	765
Ile	Glu	Arg	Val	Tyr	Arg	Arg	Ala	Gly	Ser	Gln	Lys	Leu	Trp	Ser			
															770	775	780
Val	Val	Arg	Arg	Ala	Ala	Ser	Leu	Leu	Ser	Lys	Val	Val	Asp	Ser			
															785	790	795
Leu	Ala	Pro	Ser	Ile	Thr	Asn	Val	Leu	Val	Gln	Gly	Lys	Gln	Val			
															800	805	810
Thr	Leu	Gly	Ala	Phe	Gly	His	Glu	Glu	Glu	Val	Ile	Ser	Asn	Pro			
															815	820	825
Leu	Ser	Pro	Arg	Val	Ile	Gln	Asn	Ile	Ile	Tyr	Tyr	Lys	Cys	Asn			
															830	835	840
Thr	His	Asp	Glu	Arg	Glu	Ala	Val	Ile	Gln	Gln	Glu	Leu	Val	Ile			
															845	850	855
His	Ile	Gly	Trp	Ile	Ile	Ser	Asn	Asn	Pro	Glu	Leu	Phe	Ser	Gly			
															860	865	870
Met	Leu	Lys	Ile	Arg	Ile	Gly	Trp	Ile	Ile	His	Ala	Met	Glu	Tyr			
															875	880	885
Glu	Leu	Gln	Ile	Arg	Gly	Gly	Asp	Lys	Pro	Ala	Leu	Asp	Leu	Tyr			
															890	895	900
Gln	Leu	Ser	Pro	Ser	Glu	Val	Lys	Gln	Leu	Leu	Leu	Asp	Ile	Leu			
															905	910	915
Gln	Pro	Gln	Gln	Asn	Gly	Arg	Cys	Trp	Leu	Asn	Arg	Arg	Gln	Ile			
															920	925	930
Asp	Gly	Ser	Leu	Asn	Arg	Thr	Pro	Thr	Gly	Phe	Tyr	Asp	Arg	Val			
															935	940	945
Trp	Gln	Ile	Leu	Glu	Arg	Thr	Pro	Asn	Gly	Ile	Ile	Val	Ala	Gly			
															950	955	960
Lys	His	Leu	Pro	Gln	Gln	Pro	Thr	Leu	Ser	Asp	Met	Thr	Met	Tyr			
															965	970	975
Glu	Met	Asn	Phe	Ser	Leu	Leu	Val	Glu	Asp	Thr	Leu	Gly	Asn	Ile			
															980	985	990
Asp	Gln	Pro	Gln	Tyr	Arg	Gln	Ile	Val	Val	Glu	Leu	Leu	Met	Val			
															995	1000	1005
Val	Ser	Ile	Val	Leu	Glu	Arg	Asn	Pro	Glu	Leu	Glu	Phe	Gln	Asp			
															1010	1015	1020
Lys	Val	Asp	Leu	Asp	Arg	Leu	Val	Lys	Glu	Ala	Phe	Asn	Glu	Phe			
															1025	1030	1035
Gln	Lys	Asp	Gln	Ser	Arg	Leu	Lys	Glu	Ile	Glu	Lys	Gln	Asp	Asp			
															1040	1045	1050
Met	Thr	Ser	Phe	Tyr	Asn	Thr	Pro	Pro	Leu	Gly	Lys	Arg	Gly	Thr			
															1055	1060	1065
Cys	Ser	Tyr	Leu	Thr	Lys	Ala	Val	Met	Asn	Leu	Leu	Leu	Glu	Gly			
															1070	1075	1080
Glu	Val	Lys	Pro	Asn	Asn	Asp	Asp	Pro	Cys	Leu	Ile	Ser					
															1085	1090	

<210> 12

<211> 1009

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: '7503117CD1

<400> 12

Met	Glu	Val	Val	Gly	Asp	Phe	Glu	Tyr	Ser	Lys	Arg	Asp	Leu	Val			
1				5					10					15			
Gly	His	Gly	Ala	Phe	Ala	Val	Val	Phe	Arg	Gly	Arg	His	Arg	Gln			

				20					25					30
Lys	Thr	Asp	Trp	Glu	Val	Ala	Ile	Lys	Ser	Ile	Asn	Lys	Lys	Asn
				35					40					45
Leu	Ser	Lys	Ser	Gln	Ile	Leu	Leu	Gly	Lys	Glu	Ile	Lys	Ile	Leu
				50					55					60
Lys	Glu	Leu	Gln	His	Glu	Asn	Ile	Val	Ala	Leu	Tyr	Asp	Val	Gln
				65					70					75
Glu	Leu	Pro	Asn	Ser	Val	Phe	Leu	Val	Met	Glu	Tyr	Cys	Asn	Gly
				80					85					90
Gly	Asp	Leu	Ala	Asp	Tyr	Leu	Gln	Ala	Lys	Gly	Thr	Leu	Ser	Glu
				95					100					105
Asp	Thr	Ile	Arg	Val	Phe	Leu	His	Gln	Ile	Ala	Ala	Ala	Met	Arg
				110					115					120
Ile	Leu	His	Ser	Lys	Gly	Ile	Ile	His	Arg	Asp	Leu	Lys	Pro	Gln
				125					130					135
Asn	Ile	Leu	Leu	Ser	Tyr	Ala	Asn	Arg	Arg	Lys	Ser	Ser	Val	Ser
				140					145					150
Gly	Ile	Arg	Ile	Lys	Ile	Ala	Asp	Phe	Gly	Phe	Ala	Arg	Tyr	Leu
				155					160					165
His	Ser	Asn	Met	Met	Ala	Ala	Thr	Leu	Cys	Gly	Ser	Pro	Met	Tyr
				170					175					180
Met	Ala	Pro	Glu	Val	Ile	Met	Ser	Gln	His	Tyr	Asp	Ala	Lys	Ala
				185					190					195
Asp	Leu	Trp	Ser	Ile	Gly	Thr	Val	Ile	Tyr	Gln	Cys	Leu	Val	Gly
				200					205					210
Lys	Pro	Pro	Phe	Gln	Ala	Asn	Ser	Pro	Gln	Asp	Leu	Arg	Met	Phe
				215					220					225
Tyr	Glu	Lys	Asn	Arg	Ser	Leu	Met	Pro	Ser	Ile	Pro	Arg	Glu	Thr
				230					235					240
Ser	Pro	Tyr	Leu	Ala	Asn	Leu	Leu	Leu	Gly	Leu	Leu	Gln	Arg	Asn
				245					250					255
Gln	Lys	Asp	Arg	Met	Asp	Phe	Glu	Ala	Phe	Phe	Ser	His	Pro	Phe
				260					265					270
Leu	Glu	Gln	Gly	Pro	Val	Lys	Lys	Ser	Cys	Pro	Val	Pro	Val	Pro
				275					280					285
Met	Tyr	Ser	Gly	Ser	Val	Ser	Gly	Ser	Ser	Cys	Gly	Ser	Ser	Pro
				290					295					300
Ser	Cys	Arg	Phe	Ala	Ser	Pro	Pro	Ser	Leu	Pro	Asp	Met	Gln	His
				305					310					315
Ile	Gln	Glu	Glu	Asn	Leu	Ser	Ser	Pro	Pro	Leu	Gly	Pro	Pro	Asn
				320					325					330
Tyr	Leu	Gln	Val	Ser	Lys	Asp	Ser	Ala	Ser	Thr	Ser	Ser	Lys	Asn
				335					340					345
Ser	Ser	Cys	Asp	Thr	Asp	Asp	Phe	Val	Leu	Val	Pro	His	Asn	Ile
				350					355					360
Ser	Ser	Asp	His	Ser	Cys	Asp	Met	Pro	Met	Gly	Thr	Ala	Gly	Arg
				365					370					375
Arg	Ala	Ser	Asn	Glu	Phe	Leu	Val	Cys	Gly	Gly	Gln	Cys	Gln	Pro
				380					385					390
Thr	Val	Ser	Pro	His	Ser	Glu	Thr	Ala	Pro	Ile	Pro	Val	Pro	Thr
				395					400					405
Gln	Ile	Arg	Asn	Tyr	Gln	Arg	Ile	Glu	Gln	Asn	Leu	Thr	Ser	Thr
				410					415					420
Ala	Ser	Ser	Gly	Thr	Asn	Val	His	Gly	Ser	Pro	Arg	Ser	Ala	Val
				425					430					435
Val	Arg	Arg	Ser	Asn	Thr	Ser	Pro	Met	Gly	Phe	Leu	Arg	Pro	Gly
				440					445					450
Ser	Cys	Ser	Pro	Val	Pro	Ala	Asp	Thr	Ala	Gln	Thr	Val	Gly	Arg
				455					460					465
Arg	Leu	Ser	Thr	Gly	Ser	Ser	Arg	Pro	Tyr	Ser	Pro	Ser	Pro	Leu
				470					475					480
Val	Gly	Thr	Ile	Pro	Glu	Gln	Phe	Ser	Gln	Cys	Cys	Cys	Gly	His
				485					490					495

Pro	Gln	Gly	His	Asp	Ser	Arg	Ser	Arg	Asn	Ser	Ser	Gly	Ser	Pro
				500					505					510
Val	Pro	Gln	Ala	Gln	Ser	Pro	Gln	Ser	Leu	Leu	Ser	Gly	Ala	Arg
				515					520					525
Leu	Gln	Ser	Ala	Pro	Thr	Leu	Thr	Asp	Ile	Tyr	Gln	Asn	Lys	Gln
				530					535					540
Lys	Leu	Arg	Lys	Gln	His	Ser	Asp	Pro	Val	Cys	Pro	Ser	His	Thr
				545					550					555
Gly	Ala	Gly	Tyr	Ser	Tyr	Ser	Pro	Gln	Pro	Ser	Arg	Pro	Gly	Ser
				560					565					570
Leu	Gly	Thr	Ser	Pro	Thr	Lys	His	Leu	Gly	Ser	Ser	Pro	Arg	Ser
				575					580					585
Ser	Asp	Trp	Phe	Phe	Lys	Thr	Pro	Leu	Pro	Thr	Ile	Ile	Gly	Ser
				590					595					600
Pro	Thr	Lys	Thr	Thr	Ala	Pro	Phe	Lys	Ile	Pro	Lys	Thr	Gln	Ala
				605					610					615
Ser	Ser	Asn	Leu	Leu	Ala	Leu	Val	Thr	Arg	His	Gly	Pro	Ala	Glu
				620					625					630
Glu	Gln	Ser	Lys	Asp	Gly	Asn	Glu	Pro	Arg	Glu	Cys	Ala	His	Cys
				635					640					645
Leu	Leu	Val	Gln	Gly	Ser	Glu	Arg	Gln	Arg	Ala	Glu	Gln	Gln	Ser
				650					655					660
Lys	Ala	Val	Phe	Gly	Arg	Ser	Val	Ser	Thr	Gly	Lys	Leu	Ser	Asp
				665					670					675
Gln	Gln	Gly	Lys	Thr	Pro	Ile	Cys	Arg	His	Gln	Gly	Ser	Thr	Asp
				680					685					690
Ser	Leu	Asn	Thr	Glu	Arg	Pro	Met	Asp	Ile	Gly	Ser	Pro	Pro	His
				695					700					705
Ser	Ala	Ala	Ala	Pro	Thr	Cys	Thr	His	Met	Phe	Leu	Arg	Thr	Arg
				710					715					720
Thr	Thr	Ser	Val	Gly	Pro	Ser	Asn	Ser	Gly	Gly	Ser	Leu	Cys	Ala
				725					730					735
Met	Ser	Gly	Arg	Val	Cys	Val	Gly	Ser	Pro	Pro	Gly	Pro	Gly	Phe
				740					745					750
Gly	Ser	Ser	Pro	Pro	Gly	Ala	Glu	Ala	Ala	Pro	Ser	Leu	Arg	Tyr
				755					760					765
Val	Pro	Tyr	Gly	Ala	Ser	Pro	Pro	Ser	Leu	Glu	Gly	Leu	Ile	Thr
				770					775					780
Phe	Glu	Ala	Pro	Glu	Leu	Pro	Glu	Glu	Thr	Leu	Met	Glu	Arg	Glu
				785					790					795
His	Thr	Asp	Thr	Leu	Arg	His	Leu	Asn	Val	Met	Leu	Met	Phe	Thr
				800					805					810
Glu	Cys	Val	Leu	Asp	Leu	Thr	Ala	Met	Arg	Gly	Gly	Asn	Pro	Glu
				815					820					825
Leu	Cys	Thr	Ser	Ala	Val	Ser	Leu	Tyr	Gln	Ile	Gln	Glu	Ser	Val
				830					835					840
Val	Val	Asp	Gln	Ile	Ser	Gln	Leu	Ser	Lys	Asp	Trp	Gly	Arg	Val
				845					850					855
Glu	Gln	Leu	Val	Leu	Tyr	Met	Lys	Ala	Ala	Gln	Leu	Leu	Ala	Ala
				860					865					870
Ser	Leu	His	Leu	Ala	Lys	Ala	Gln	Ile	Lys	Ser	Gly	Lys	Leu	Ser
				875					880					885
Pro	Ser	Thr	Ala	Val	Lys	Gln	Val	Val	Lys	Asn	Leu	Asn	Glu	Arg
				890					895					900
Tyr	Lys	Phe	Cys	Ile	Thr	Met	Cys	Lys	Lys	Leu	Thr	Glu	Lys	Leu
				905					910					915
Asn	Arg	Phe	Phe	Ser	Asp	Lys	Gln	Arg	Phe	Ile	Asp	Glu	Ile	Asn
				920					925					930
Ser	Val	Thr	Ala	Glu	Lys	Leu	Ile	Tyr	Asn	Cys	Ala	Val	Glu	Met
				935					940					945
Val	Gln	Ser	Ala	Ala	Leu	Asp	Glu	Met	Phe	Gln	Gln	Thr	Glu	Asp
				950					955					960
Ile	Val	Tyr	Arg	Tyr	His	Lys	Ala	Ala	Leu	Leu	Leu	Glu	Gly	Leu

	965		970		975
Ser Arg Ile Leu	Gln Asp Pro Ala Asp Ile	Glu Asn Val His	Lys		
	980		985		990
Tyr Lys Cys Ser	Ile Glu Arg Arg Leu Ser	Ala Leu Cys His	Ser		
	995		1000		1005
Thr Ala Thr Val					

<210> 13

<211> 405

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7506911CD1

<400> 13

Met Lys Leu Ile	Asn Gly Lys Lys Gln Arg	Thr Thr Ser Asp	Glu
1	5	10	15
Glu Ser Gly Ile	Ile Lys Ala Glu Asp Leu	Thr Pro Val Ile	Pro
	20	25	30
Pro Leu Trp Glu	Ala Glu Ala Gly Gly Ser	Arg Ala Phe Pro	Trp
	35	40	45
Phe Gly Met Asp	Ile Gly Gly Thr Leu Val	Lys Leu Val Tyr	Phe
	50	55	60
Glu Pro Lys Asp	Ile Thr Ala Glu Glu Glu	Gln Glu Glu Val	Glu
	65	70	75
Asn Leu Lys Ser	Ile Arg Lys Tyr Leu Thr	Ser Asn Thr Ala	Tyr
	80	85	90
Gly Lys Thr Gly	Ile Arg Asp Val His Leu	Glu Leu Lys Asn	Leu
	95	100	105
Thr Met Cys Gly	Arg Lys Gly Asn Leu His	Phe Ile Arg Phe	Pro
	110	115	120
Ser Cys Ala Met	His Arg Phe Ile Gln Met	Gly Ser Glu Lys	Asn
	125	130	135
Phe Ser Ser Leu	His Thr Thr Leu Cys Ala	Thr Gly Gly Gly	Ala
	140	145	150
Phe Lys Phe Glu	Glu Asp Phe Arg Met Ile	Ala Asp Leu Gln	Leu
	155	160	165
His Lys Leu Asp	Glu Leu Asp Cys Leu Ile	Gln Gly Leu Leu	Tyr
	170	175	180
Val Asp Ser Val	Gly Phe Asn Gly Lys Pro	Glu Cys Tyr Tyr	Phe
	185	190	195
Glu Asn Pro Thr	Asn Pro Glu Leu Cys Gln	Lys Lys Pro Tyr	Cys
	200	205	210
Leu Asp Asn Pro	Tyr Pro Met Leu Leu Val	Asn Met Gly Ser	Gly
	215	220	225
Val Ser Ile Leu	Ala Val Tyr Ser Lys Asp	Asn Tyr Lys Arg	Val
	230	235	240
Thr Gly Thr Ser	Leu Gly Gly Gly Thr Phe	Leu Gly Leu Cys	Cys
	245	250	255
Leu Leu Thr Gly	Cys Glu Thr Phe Glu Glu	Ala Leu Glu Met	Ala
	260	265	270
Ala Lys Gly Asp	Ser Thr Asn Val Asp Lys	Leu Val Lys Asp	Ile
	275	280	285
Tyr Gly Gly Asp	Tyr Glu Arg Phe Gly Leu	Gln Gly Ser Ala	Val
	290	295	300
Ala Ser Ser Phe	Gly Asn Met Met Ser Lys	Glu Lys Arg Asp	Ser
	305	310	315
Ile Ser Lys Glu	Asp Leu Ala Arg Ala Thr	Leu Val Thr Ile	Thr
	320	325	330
Asn Asn Ile Gly	Ser Ile Ala Arg Met Cys	Ala Leu Asn Glu	Asn

Ile Asp Arg Val	335	Val Phe Val Gly Asn	340	Phe Leu Arg Ile Asn Met	345
Val Ser Met Lys	350	Leu Leu Ala Tyr Ala	355	Met Asp Phe Trp Ser Lys	360
Gly Gln Leu Lys	365	Ala Leu Phe Leu Glu	370	His Glu Gly Tyr Phe Gly	375
Ala Val Gly Ala	380	Leu Leu Glu Leu Phe	385	Lys Met Thr Asp Asp Lys	390
	395		400		405

<210> 14
 <211> 226
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7510809CD1

<400> 14

Met Ala Gly Tyr Lys	Pro Val Ala Ile	Gln Thr Tyr Pro Ile	Leu
1	5	10	15
Gly Glu Lys Ile Thr	Gln Asp Thr Leu	Tyr Trp Asn Asn Tyr	Lys
	20	25	30
Thr Pro Val Gln Ile	Lys Glu Phe Gly	Ala Val Ser Lys Val	Asp
	35	40	45
Phe Ser Pro Gln Pro	Pro Tyr Asn Tyr	Ala Val Thr Ala Ser	Ser
	50	55	60
Arg Ile His Ile Tyr	Gly Arg Tyr Ser	Gln Glu Pro Ile Lys	Thr
	65	70	75
Phe Ser Arg Phe Lys	Asp Thr Ala Tyr	Cys Ala Thr Phe Arg	Gln
	80	85	90
Asp Gly Arg Leu Leu	Val Ala Gly Ser	Glu Asp Gly Gly Val	Gln
	95	100	105
Leu Phe Asp Ile Ser	Gly Arg Ala Pro	Leu Arg Gln Phe Glu	Gly
	110	115	120
His Thr Lys Ala Val	His Thr Val Asp	Phe Thr Ala Asp Lys	Tyr
	125	130	135
His Val Val Ser Gly	Ala Asp Asp Tyr	Thr Val Lys Leu Trp	Asp
	140	145	150
Ile Pro Asn Ser Lys	Glu Ile Leu Thr	Phe Lys Glu His Ser	Asp
	155	160	165
Tyr Val Arg Cys Gly	Cys Ala Ser Lys	Leu Asn Pro Asp Leu	Phe
	170	175	180
Ile Thr Gly Ser Tyr	Asp His Thr Val	Lys Met Phe Asp Ala	Arg
	185	190	195
Thr Ser Glu Ser Val	Leu Ser Val Glu	His Gly Gln Pro Val	Glu
	200	205	210
Ser Val Leu Leu Phe	Pro Ser Gly Gly	Leu Leu Val Ser Ala	Gly
	215	220	225

Arg

<210> 15
 <211> 4344
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 2537210CB1

<400> 15

```

cgcagttcgg ggcaggtctg ggcaggtctg ggcaggtctg ggcaggtctg ggcaggtctg 60
ggcaggtctg ggcaggtctg ggcaggtctg ggcaggtctg ggcaggtctg ggcaggtctg 120
gcagacgggtc tggcaggtc ccccatcttga ggcaggtcag ggcaggtcag ggcaggtcag 180
tggcaggtc gtcagcatcc actccttcca ggcaggtcag ggcaggtcag ggcaggtcag 240
gtccatcatc cccaacaagg tggccctgt tgtgatcacg tacaactgca aggaggagtt 300
ccagatccat gatgagctgc tcaaggctca ttacacgttg ggcaggtctc cggacaacac 360
ccctgagcac tacctggtgc aaggccgcta ctctctggtg cgggatgtca ctgagaagat 420
ggatgtgctg ggcacccgtg gaagctgtgg ggccccaac ttccggcagg tgcaggtgtg 480
gctcactgtg ttccggcatgg gacagcccag cctctcaggg ttccaggcggg tcctccagaa 540
actccagaag gacggacata gggagtgtgt catcttctgt gtgcccaggg aacctgtgct 600
tttccctgct gcagatgagg actttgtgtc ctacacacct cgagacaagc agaaccttca 660
tgagaacctc cagggccttg gacccggggt ccgggtggag agcctggagc tggccatccg 720
gaaagagatc cacgactttg cccagctgag cgagaacaca taccatgtgt accataacac 780
cgaggacctg tggggggagc cccatgctgt ggccatccat ggtgaggacg acttgcattg 840
gacggaggag gtgtacaagc ggcctctctt cctgcagccc acctacaggt accaccgct 900
gcccctgccc gagcaaggga gtcccctgga ggccagttg gacgcctttg tcagtgttct 960
ccgggagacc cccagcctgc tgcagctccg tgatgccac gggcctcccc cagccctcgt 1020
cttcagctgc cagatgggag tgggcaggac caacctgggc atggctcctg gcacctcat 1080
cctgcttcac cgcagtggga ccacctcca gccagaggct gcccacagc agccaagcc 1140
cctgcttatg gagcagttcc aggtgatcca gagctttctc cgcaggtgc cccagggaag 1200
gaggatgggt gaagaggtgg acagagccat cactgctgtg gccaggtgc atgacctgaa 1260
agaagtgggt ttggaaaacc agaagaagt agaaggtatc cgaccggaga gccagacca 1320
gggaagcggc agccgacaca gcgtctggca gagggcgtg tggagcctgg agcgatactt 1380
ctacctgatc ctgtttaact actaccttca tgagcagtac ccgctggcct ttgcccctcag 1440
tttcagccgc tggctgtgtg cccaccctga gctgtaccgc ctgcccgtga cgtgagctc 1500
agcaggccct gtggctccga gggacctcat cgccaggggc tccctacggg aggacgatct 1560
ggtctccccg gacgcgctca gcaactgtcag agagatggat gtggccaact tccggcgggt 1620
gcccgcagtg cccatctacg gcacggccca gccagcgc aaggccctgg ggagcaccct 1680
ggcctacctg acggacgcca agaggaggct gcggaagggt gtctgggtga gccttcggga 1740
ggaggccgtg ttggagtgtg acgggcacac ctacagcctg cgggtggcctg ggcccctgt 1800
ggctcctgac cagctggaga cctggaggc ccagctgaag gccatctaa gcgagcctcc 1860
cccaggcaag gagggcccc tgcctacag gtccagacc tgccttacca tgcaggaggt 1920
cttcagccag caccgcaggg cctgtcctgg cctcacctac caccgcctcc ccatgccgga 1980
cttctgtgcc ccccgagagg aggactttga ccagctgctg gaggccctgc gggccgccc 2040
ctccaaggac ccaggcactg gcttctgtgt cagctgcctc agcggccagg gccgtaccac 2100
aactgcgatg gtgggtggctg tcttggcctt ctggcacatc caaggcttcc ccgaggtggg 2160
tgaggaggag ctctgtgagt tgcctgatgc caagttcact aagggtgaat ttcaggtagt 2220
aatgaagggt gtgcagctgc taccgatgg gcaccgtgtg aagaaggagg tggacgcagc 2280
gctggacact gtcagcgaga ccatgacgcc catgcactac cacctgcggg agatcatcat 2340
ctgcacctac cgccaggcga aggcagcgaa agaggcgcag gaaatgcgga ggctgcagct 2400
gcccagcctg cagtacttgg agcgtatgt ctgctgatt ctcttcaacg cgtacctcca 2460
cctggagaag gccgactcct ggcagaggcc cttcagcacc tggatgcagg aggtggcatc 2520
gaaggctggc atctacgaga tcttaacga gctgggcttc cccgagctgg agagcgggga 2580
ggaccagccc ttctccaggc tgcgctaccg gtggcaggag cagagctgca gcctcgagcc 2640
ctctgcccc gaggacttgc tgtagggggc cttactccct gtccccccac ccacagggcc 2700
ccacgcaggc ctgggggtgtc tgaggtgtct ttggctggga ggcggcctga ggggtgctgg 2760
ccttgaaatg attccccac ttctggaga gactgagcgg agttgggagc ctttttagaa 2820
agaacttttt ataggacagg gagacagcac agccatccct tgcaaaccac caaggtgtgt 2880
ggctgacctc caggaggagg cactcactgg agtgctcaca aggtgcacac tgctgtgtgt 2940
accttgacga caggccggcg ttccagcctc aaggggctca ctccccagc tgccaaacac 3000
tgtggatctc tctgtcctct tctccccctc ctccagattg cctggcagcc cctgtcatgg 3060
agcagacccg gccactggta gctccccact tcttactcc tgctgctctg ccattgccgc 3120
tccccttctt gctgcccagg cactgcccct gggcgtcttg cagcctgagg tgggtggagg 3180
ggacagtgtt ctggatagat ctattatgtg aaaggcagct tcacccagtt ttctggactc 3240
tcatgcccc atctccgacc tgggagactt caggaatgac aacctacca gcctgggtggg 3300
gctggcagga tgggtggagg ttctcaagga gctggagact tcaggagacc cctctcatgg 3360
ggaggaaaga gcttccaggg ggcgaacgca gcacagagga agaggcctgc tccacttgtc 3420
tgggaacctg ggcaggaggc acagagggaag ccaaggcctg gagctgcagg tccccgggca 3480
tctctctctg tcccggcagc ccaggatggc ctgggtgccc cacctgctgc agcaggagcc 3540
ccaaggagtg ctgctgagg gtggttgcgt gggtggctct catggacagt gaggtgtgca 3600
agggtgactt gagggtgggt ggaggggatc acctgggttc caggccatcc ttgctgagca 3660
tctttgagcc tgccttccgg tgggagcaga aaaggccaga ccctgctgag ttagaggctg 3720

```

```

ctgggatcca ctgtttccac acagcgggaa ggctgctggg aacaggtggc agagaagtgc 3780
catgtttgcg ttgagccttg cagctcttcc agctggggac tgggtgcttg tgaaaccag 3840
gagctgaaca gtgaggaggc tgtccacctt gcttggctca ctgggaccag gaaagcctgt 3900
ctttgggttag gctcgtgtac ttctgcagga aaaaaaaaaa aggatgtgtc attggtcattg 3960
atatttgaaa aggggaggag gccgaagttg ttcccattta tccagtattg gaaaatattt 4020
gaccccttg gctgaattct tttgcagaac tactgtgtgt ctgttacta ccttttcagg 4080
tttattgttt ttatttttgc atgaattaag acgttttaat ttctttgcag acaaggctta 4140
gatgcggagt cagagatggg actgaatggg gagggatcct ttgtgttctc atggttggct 4200
ctgactttca gctgtgttgg gaccactggc tgatcacatc acctctctgc ctcagtttcc 4260
ccatctgtaa aatgggagaa taatacttgc ctacctacct cacaggggtg ttgtgaggat 4320
tcatttgtga tttttttttt tttt 4344

```

<210> 16

<211> 5621

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 112535CB1

<400> 16

```

ccgcttcgcc agacagccag cggccggcgg caggccgggc catgagcggc aggggcccgg 60
ccgggcctcg ctgaccctgg ctccgcgcgg cagcttcccc agtttccgct ccggtctctc 120
ggcatgagag tccgcccggg cccggggctg cggctgcccc agaccgcgcg cacgctggcg 180
cgctccgggc ccgcgaggac gcggtgctga tactgcgcc gactgcgcc gcccgccgt 240
ccgctgtgtg ccccgggggc gcggccatgg aggtgggtgg tgacttcgag tacagcaaga 300
gggatctcgt gggacacggg gccttcgccg tggctctccg ggggcggcac cgccagaaaa 360
ctgattggga ggtagctatt aaaagtatta ataaaaagaa cttgtcaaaa tcacaaatac 420
tgcttggaiaa ggaaattaaa atcttaaaagg aacttcagca tgaaaatatt gtagcactct 480
atgatgttca ggaattaccc aactctgtct ttttgggtgat ggagtattgc aatggtggag 540
acctcgcaga ttatttgcaa gcgaaaggga ctctcagtga agacacgatc agagtgtttc 600
tgcatcagat tgctgctgcc atgcgaatcc tgcacagcaa aggaatcatc cacagagatc 660
tcaaaccaca gaacatcttg ctgtcctatg ccaatcgcag aaaatcaagt gtcagtggta 720
ttcgcatcaa aatagcggat tttgggtttg ctcggtacct acatagtaac atgatggctg 780
caacactgtg tggatccccg atgtacatgg ctcttgaggt tattatgtct caacattatg 840
atgctaaggc tgacttgttg agcataggaa cagtgatata ccaatgccta gttggaaaac 900
caccttttca ggccaatagt cctcaagact taaggatgtt ttatgaaaaa aacaggagct 960
taatgcctag tattcccaga gaaacatcac cttatttggc taatctcctt ttgggtttgc 1020
ttcagagaaa ccaaaaagat agaattggact ttgaagcgtt ttttagccat ccttttcttg 1080
agcaaggctc agtaaaaaaa tcttgcccag ttccagtgcc catgtattct ggttctgtct 1140
ctggaagctc ctgtggcagc tctccatctt gtcgttttgc ttctccacca tcccttccag 1200
atatgcagca tattcaggaa gaaaacttat cttccccacc attgggtcct cccaactatc 1260
tacaagtttc caaagattct gccagtacta gtagcaagaa ctcttcttgt gacacggatg 1320
actttgtttt ggtgccacac aacatctcgt cagaccactc atgtgatatg ccaatgggga 1380
ctgctggcag acgtgcttca aatgaattct tgggtgtgtg agggcagtgt cagcctactg 1440
tgtcacctca cagcgaacaa gcaccaatc cagttcctac tcaaataagg aattatcagc 1500
gcatagagca gaatcttaca tctactgcc a gctcaggcac aaatgtacat ggttctccaa 1560
gatctgcagt ggtacgaagg tccaacacca gcccctatgg cttcctccgg ccgggatcat 1620
gctccccagt accagcagac acagcacaga cagttggacg aaggctctcc actgggtctt 1680
ctaggcctta ctacacttcc cctttgggtg gtaccattcc tgagcaatc agtcagtgt 1740
gctgtgggca tcttcagggc catgactcca ggagtagaaa ctctcagggt tctccagtgc 1800
cacaagctca gtccccacag tctctcttat cgggtgctag actgcagagc gccccaccc 1860
tactgacat ctatcagaac aagcagaagc tcagaaaaca gcactctgac ccggtgtgcc 1920
catcccatac tggggctggg tacagctact cgcctcagcc cagtcggcct ggcagccttg 1980
gaacttctcc caccaagcac ttgggggtcct ctccacggag ttctgactgg ttctttaaaa 2040
ctcctttgcc aacaatcatt ggctctccta ctaagaccac agctcctttc aaaatcccta 2100
aaactcaagc atcttccaac ctgttagcct tgggtactcg tcatgggcct gctgaagaac 2160
agtgcgaaaga tgggaatgag ccacgggaat gtgcccatgt cctcttagtg caaggaagt 2220
agaggcagcg ggccgagcag cagagcaagg cagtgttttg cagatctgtc agtaccggga 2280
agttatcaga tcaacaagga aagactccta tatgtcgaca tcagggcagc acagacagtt 2340
taaatacaga acgaccaatg gatatagtct cggcaggagc ctgtgggtgg gttctggcac 2400
ctcctgcagg tacagcagca agttccaagg ctgtcctctt cactgtaggg tctcctccac 2460

```

```

acagtgcggc agccccact tgtaccacaca tgttccttcg aacaagaaca acctcagtgg 2520
ggcccagcaa ctccgggggc tctctttgtg ccatgagtgg ccgcgtgtgc gtgggggtccc 2580
cgccctggccc aggccttcggc tcttccccctc caggagcaga ggcagctccc agcctgagat 2640
acgtgcctta cgggtgcttca cccccagcc tagaggggct catcaccttt gaagcccctg 2700
aactgccgga ggagacgctg atggagcggg aacacacaga caccttacgc catctgaatg 2760
tgatgctgat gttcactgag tgtgtgctgg acctgacagc catgagggga ggaaaccctg 2820
agctgtgcac atctgctgtg tccttgtacc agatccagga gagtgtggtg gtggaccaga 2880
tcagtcaagt gagcaaagac tggggggcggg tggagcagct ggtgttgtac atgaaagcag 2940
cacagctgct tgcggcttct ctgcatcttg ccaaagccca gatcaagtcc gggaaactga 3000
gcccattcac agctgtgaaa caagtgtgca agaacttgaa cgaacgatat aaattctgca 3060
tcaccatgtg caagaaactt acagaaaagc tgaatcgatt cttctctgac aaacagaggt 3120
ttattgatga aatcaacagt gtgactgcag agaaactcat ctataattgt gctgtagaaa 3180
tgggttcagtc tgcagccctg gatgagatgt ttcagcagac cgaagatatt gtttatcgct 3240
atcataaggc agcccttctt ttggaaggcc taagtaggat tctacaggac cctgcagata 3300
ttgaaaatgt gcataaatat aaatgtagta ttgagagaag actgtcggcg ctctgccata 3360
gcaccgcaac cgtgtgagca gcaggtcat cccgtggacc ggtggtggga acgtgaggtg 3420
atgccttttg gattacagct tgagttctgt caccatcc ccaggaaact gtagcttctt 3480
aactggtgac taccaaagaa caagcagtga tttgaaaaag gaaaaacaat ccaaaaacta 3540
catatttgta ggaaatctgc cttattggag aaaatcacc tttccctttt tctttgtaga 3600
agcaggagca agagtgtttg gctcccagtt tggacttgg gaataaatgt accttagaac 3660
taggataatc ggtacagtta ttcttaaga taattaaaa tgaaacaaag tgagtgtctg 3720
tcactgggtt catcagagca gtgtgtgaaa ttccatgtgt ttgctgaggt gtaaaggtaa 3780
atgtattcac ccctcatcca ggcagtttga tatttggagt aagtttgttt aaatctgagc 3840
atgcatcttt aaacagctca ggaagaaata gcttaagaag aagtgaaca tggatcttgg 3900
aagaaatttt gaaatcttca atttgatcct aatatggata catgttaatc ttccaaatc 3960
tttcatattg cactaattta ttaaaacaac tgtgtattgg attttgtaat ttaactaagg 4020
cacaatggac ttgtttaaaa tattttactt gattgtatac atagaccctt tccagaattc 4080
acatgtaatc tccagtgaac ttttaagtgg ttaaaacttg tattcatgtg aacctttgca 4140
catttttttt tttttacttc tttatctaca cctacagatt ttctcagtaa tgtttttgtt 4200
agcttttggg tccatttttt attgtgcatg cagaatgtac attgatgcct gtgaccttag 4260
gtttattaaa ggctagggtt atttgggcag tattagaaac aaaatcatgg atcaagagat 4320
actcttgata atttgaatag ggccaaaaca aagttggtga cctaaaggct tgttagtgat 4380
gtggagttcc tacatgcagt gagtggaaaa tgaagtctgt tttctcttag gaaaatgggc 4440
agctgtcttc tgcctaattg gtatttttca tgttaattct gacagttcac caaatagcta 4500
gtcatggaga atgcaggcag ttaacttaat atccctccag gaatgggtcc tacgttgtgt 4560
attatttggg ttcttttact tacctgcttg aatacttgaa taaaccattc accaatttta 4620
atccttttat tttaatcctt ttacataaaa taatctgaac tctttgacaa attgcacaga 4680
gctctttggc attaatctaa ttttaatgta ctgataaaaa caaacatggg tgccttttac 4740
tttgacaaag taatgtaatt tttaccttat ttatctgtat gaaattccag tagttaattt 4800
gaacatttat ttatatgacg tttgtatttt taggtcttta atacagtgtt tctacctctc 4860
atttgtaact gcatgcatta ttcttgaaac taggtaaaac tcactgaatt gttgtgtaat 4920
agccttttta ttattgcctg taaaaatgta tattaaggta aaataaaaact gacaaagtgt 4980
ttctaggggt tagctgggta catattaagt ggcttggtga gccagggtact tccttagtga 5040
gttttagagac ttggccatga atatcctttg tcctgcccc aagatttagat cttggctact 5100
gtcatgcagg ctccaggaa catagactgt tttacctcca caacctatt tgttattagt 5160
gatactttat tttatataat attttttatt cacagtgaat tttcattcat gttctttcag 5220
ttatcacctg tgttatctca gttgtagggt tattctatcc tctcctcttc ctctccatt 5280
tcttttttaa cacaggatga aacagggttca gagaggggaa gtgattggcc taaagtcagg 5340
aactaggcaa gtggtcaagc catgctttgt gactttcaag ttaattcttc ttgttcttgt 5400
atattaaagg tcttggggta gatggtgtgt gtgaaacagt gaagtctcaa cagcagaaaa 5460
gaacaaaatg taaattcatg aataatggtt ctggttatac ttccattatc aaggctaatt 5520
aagagatttt gccttgagta tagcaataat aaacaaatgc tttatgtttc cctgaaaaaa 5580
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa a 5621

```

<210> 17

<211> 1581

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 72063274CB1

<400> 17

```

gccgagacgg ctacatggac gccactatcg ccccgacccg tatccccccc gagatgcccc 60
agtacgggga ggagaaccac atcttcgagt tgatgcagaa catgctggag caactcctga 120
tccaccagcc cgaagatccc atccccctca tgatccagca cttgcataga gacaacgaca 180
atgtgcccag gattgtaata ttaggtccac ccgcctcagg gaaaacaaca atagcaatgt 240
ggctctgcaa acatctgaac agcagtctcc tcaccctgga gaacctgac tttaatgagt 300
tttcctatac ggccaccgaa gccagaaggc tttatctgca aaggaagaca gttcccagcg 360
cgctgctcgt ccagctgatt caggaacgcc tggctgaaga ggattgcac aagcagggct 420
ggattctgga tggcatccct gagacgcgtg agcaggctct gaggatccag accctgggga 480
tcacaccag acacgtcatt gtgctgagtg ctccagacac ggtcctgac gagagaaact 540
tggggaagag aatcgacctt caaactggag agatttatca caccaccttt gactggccac 600
ccgaatctga aatccagaac cgtctcatgg tgccagagga catctcagag ctggagacgg 660
ctcagaaact gctggagtat cataggaaca tcgtcagggt cattccctcc taccctaaaa 720
tcctcaaagt catcagtgtg gaccagccat gtgtggacgt cttctaccag gctctgacct 780
atgtccaaag caaccatcgt actaatgccc cgttcacccc gaggggtgtg ctgctcgggc 840
ctgtgggcag tgggaaaagt ctgcaggccg ccctcctggc ccagaaatac aggcttgtca 900
atgtctgctg tgggcaactg ctgaaagagg ctgtggcaga taggaccacg tttggcgagc 960
tcatccagcc cttctttgaa aaggagatgg cagttcctga cagcctcctc atgaagggtg 1020
tgagccagcg cctggaccag caggactgca tccagaaagg ctgggtgcta cacggcgtcc 1080
cgcgggacct cgaccaggca cacctgctga accgcctggg ctacaatccc aacagggtgt 1140
ttttcctgaa tgtgccattt gattccatca tggagcggct gactctgaga agaattgatc 1200
cagtcactgg ggaaaggtac cacctcatgt acaagccacc tcccaccatg gagatccagg 1260
ctcgccctct gcagaaccca aaggatgctg aagagcaggt caagctgaaa atggacctgt 1320
tctacaggaa ctcagctgac ttggagcagt tgtatgggtc ggccatcacc ctcaatgggg 1380
accaggaccc atacacagtc ttcgaataca tcgagagtgg gatcattaat cccctgcccc 1440
agaaaatccc ctgatgggtt cagagccagg agcgtgccc cagggaagaa gttaatcccc 1500
tgccccagc cccccagcct cggcacagct ccctaaaaa gccataaag cctgctggat 1560
acagaaaaaa aaaaaaagg g

```

<210> 18

<211> 3156

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 5013673CB1

<400> 18

```

cggcgctaca gcaagttttt tgacctccag atgcagatgt tggacaaatt tcccatggga 60
agggaggaca gaaggacccc aagcagcgga tcatccctt tctgccaggt aagattctct 120
tcagacgaag ccacatccgg gacgtggctg tcaaacgcct gataccaatt gatgaatact 180
gtaaggccct catccagctg cccccctaca tctctcagtg tgatgaggtg ctgcagttct 240
ttgagacaag acctgaggac ctgaatcccc ccaaagagga gcacattggg aaaaagaaat 300
ctgggggtga ccaaacctca gtggacccca tggctcctgga gcagtatgtg gtggtagcca 360
actaccagaa gcaggagagt tcggagatca gcctcagcgt ggggcaggtg gtggacatca 420
tcgagaagaa tgagtcaggt tgggtggttcg tcagcactgc cgaggagcaa ggctgggtcc 480
ctgcaacgtg cctcgaaggc caggatgggg tgccaggatga gttttctctg cagcctgaag 540
aagaggagaa gtacacagtc atctacccgt acacagctcg ggaccaggat gaaatgaacc 600
tggagagagg ggctgtggtg gaggtcatcc agaaaaacct ggaaggctgg tggagatca 660
ggtaccaggg caaagaaggc tgggcccccg cctcctacct aaagaagaac agtggggagc 720
ccttgcccc gaagccaggc cctggctcac cctcccaccc ggggtgccctt gacttggatg 780
gtgtttcccg gcagcagaac gcggtgggca gggagaagga gctgctcagc agccagaggg 840
acgggcgggt tgaaggccgc ccggtgcccg acggtgacgc caagcagaga tcaccaaaag 900
tgaggcagag acccctcct cgccgggaca tgaccattcc tcgaggcctc aacctgccga 960
agccgcccac cccgccccaa gtggaggaag agtattacac catcgccgaa ttccagacaa 1020
ccatcccaga cggcatcagc ttccaggcag gcctgaaggc cgaggtgatc gagaaaaact 1080
tgagtggctg gtggtacatt cagattgaag ataaggaagg gtgggccccg gccaccttca 1140
ttgacaagta caagaagacg agcaacgcgt cgagacccaa ctttctggct cccctgcccc 1200
acgaggtgac ccagctccgg ctgggggaag cagcagcgt ggagaacaac acgggcagcg 1260
aagccacggg cccctcccg cccctgctg acgcaccgca tgggtgcatg gactcgggt 1320
tgccatggtc taaagactgg aagggcagta aggatgtcct gaggaaggca tcttcagaca 1380
tgtctgcgtc agcaggctac gaggagatct cagaccccg catggaggag aagcccagcc 1440

```

```

tccctccgcg gaaagaatcc atcatcaagt cggaggggga gctgctggag cgggagcggg 1500
agcggcagag gacggagcag ctccggggcc ccactcccaa gcctccgggc gtgattttgc 1560
cgatgatgcc agccaaacac atccctccag cccgggacag caggaggcca gagcccaaac 1620
ctgacaaaag cagactgttc cagctgaaaa atgacatggg gctggagtgt ggccacaagg 1680
tcttggccaa ggaagtgaag aagcccaacc tccggcccat ctccaaatcc aaaactgacc 1740
tgccagagga gaagccagat gccactcccc agaatccctt cttgaagtcc agacctcagg 1800
ttagggcaaa accagctcct tcccccaaaa cggagccacc tcagggcgaa gaccaagtgc 1860
acatctgcaa cctcaggagt aagctcaggc ctgccaaagc ccaagacaag tccttggttg 1920
atggggaggg cccccaggca gtagggggcc aagacgtggc cttcagccga agcttcctcc 1980
caggagaggg gcctggccgc gcccaggaca ggacgggcaa acaggatggg ctcagcccaa 2040
aagagatttc ctgcagagcc cctccgaggc cagccaagac cacagatcct gtgtctaaga 2100
gcgtagcctg tcctctccaa gaggtcccc agcagagacc tgtgggtcca ccccgagac 2160
cacctcccc aaagaaaacc tcttcgtcat ccaggccgct cccagaggtc agaggtccac 2220
agtgtgaagg ccacgaaagc agggcagctc ccaccccagg ccgtgctctc ctctccctc 2280
caaaagccaa accttttctc tccaactctt tggggggcca ggatgacacg cgaggcaaag 2340
gcagcctggg gccatggggg accggcaaga ttggagaaaa caggagagaa gcagctgcag 2400
cctctgtccc caatgccgac ggcctgaagg actctttgta tgtggccgtg gccgactttg 2460
aaggagacaa agacaccagc agcttccagg aagggacagt gtttgaagtc cgggagaaga 2520
acagcagtgg ctggtggttc tgccagggtc tgagcggagc cccttccctg gaagggtgga 2580
ttccttccaa ctatctcaga aagaagccgt agccgactcc ctttctgcct agagggccg 2640
ctggtccttg ctggctttac ccacgtatct aatacgctc ttaatttate attctccacg 2700
cagcttccaa ggcagacaga ctctggggta ctgtgacttc ttgcctcca tgggtggaga 2760
gtgagtttcg gacacctcgg gcgcccctgg gcctgatccc tcctatcaca gcatcactgg 2820
aggetcagaa cccacagcct ttgctttctg tccatgtcag catccctgcc ttaagagaac 2880
tcctcctggc caatggcatt gccacccagc agtgggacca agactctcca agacctccag 2940
gactggatcc cattgcctgg agaaactcca gcaagggtct ctcatggctt ggacatggca 3000
cagtaagggg cagccaacc agtccatgat gacttttgct ccaacttctt catgtttcta 3060
aaagcccagt ggctttattc actcctccta aattgcctgc taccagaagg aacttcatcc 3120
tgaagaaatg cattccatta ccagttccag ggaaag 3156

```

<210> 19

<211> 1567

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 5977982CB1

<220>

<221> unsure

<222> 1405

<223> a, t, c, g, or other

<400> 19

```

ctgcgatctt cccagcacag acgtttggac agagcaggct cctaaggtct ccagaatgcc 60
cgtgccagcc tcctggcccc accttcctag tcctttcctg ctgatgacgc tactgctggg 120
gggactcaca ggggtagctg gcgaggaaga gctgcagggt attcagcctg acaagtccat 180
atcagttgca gctggagagt cggccactct gcaactgcact gtgacttccc tgatccctgt 240
ggggcccatc cagtggttca gaggagctgg accaggccgg gaattaatct acaatcagaa 300
agaaggccac tccccacggg taacaactgt ttcagacctc acaaagagaa acaacatgga 360
cttttccatc cgcacagta acatcacccc agcagatgcc ggcacctact actgtgtgaa 420
gttccggaaa gggagccccg accacgtgga gtttaagtct ggagcaggca ccgagctgtc 480
tgtgcgtgcc aaacctctg ccccggtggg atcgggcccct gcggcgaggg ccacacctca 540
gcacacagtg agcttcacct gcgagtccca cggcttctca cccagagaca tccacctgaa 600
atggttcaaa aatgggaatg agctctcaga cttccagacc aacgtggacc ccgcaggaga 660
cagtgtgtcc tacagcatcc acagcacagc caagggtgtg ctgaccgcg aggacgttca 720
ctctcaagtc atctgcgagg tggcccacgt caccttgacg ggggacccctc ttcgtgggac 780
tgccaacttg tctgagacca tccgagttcc acccaccttg gaggttactc aacagcccgt 840
gagggcagag aaccagggtg atgtcacctg ccagggtgag aagtctctacc cccagagact 900
acagctgacc tggttggaga atggaaacgt gtccaggaca gaaacggcct caacccttac 960
agaaaacaag gatggtacct acaactggat gagctggctc ctggtgaatg tatctgcca 1020
cagggatgat gtgaagctca cctgccaggt ggagcatgac gggcagccag cggtcagcaa 1080

```



```

aagccatgac ctgaagggtct cagcccaccc gaaggagcag ggctcaaata ctgctcctgg 1140
cccagcactg gcttctgctg ctccacttct catagctttc ctctggggcc ccaagggtgct 1200
gctgggtggtt ggtgtctctg tcatctatgt ctactggaag cagaaggcct gactaagcac 1260
tctgaggccc tcagcagcag atgcaagtac catgcttctt gtacagtctg cagaaccaa 1320
agcaaataaa ctctctttat aaattaaaaa aaaaaaaggg ggcgcgataa gtgagctcgt 1380
cgaccggaat tattcgcgac cggtnccctgc aggcgtccag ttcctatagt gagcgggtta 1440
gaccggggga acatggcata ccgtcctggg tgaatgggtc gccacacca cgcagcggg 1500
gggggggggc gaaacgaggg gtcccaccac cccgggcgag gggggggggc gacgcgcgcc 1560
cagagag                                           1567

```

<210> 20

<211> 2212

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 6880271CB1

<220>

<221> unsure

<222> 1894

<223> a, t, c, g, or other

<400> 20

```

ggagccgcgt tccgccgaga gttggggcaga ggagcgcccg cgccccggcg gcgtcatggg 60
ccccctcccc gcgcttcaga gggcaccagc cgcggggaacc cccgggcctc ctgcgcgccg 120
agcctgagcg accctcgggt tctccggcgc cccctccctc gccctathtt ttttcctact 180
ctcgctgccg ttaccgcttc tgctctccgt tatggcaaca gagccaccat cccccctccg 240
ggtcgaggcg cggggccccc cagaaatgcg gacctcaccg gcgctcgagt ccaccctga 300
gggcaccccc cagccggcgg gcggcagact ccgcttcctc aacggctgcg tgccctctc 360
gcatcagggt gccgggcaca tgtacgggaa ggacaaagtg ggtatactgc aacatccaga 420
tggcacagtt ttgaaacagt tacaaccacc tccaaggggc ccaagagagc tggaattcta 480
taatatggtt tatgctgctg actgttttga tgggtgttct ctagagctac gaaaataatt 540
gccaaaatat tatggcatct ggtcacctcc cactgcacca aacgatttat acctaaaact 600
ggaagatgtg acccataaat ttaataagcc ctgtataatg gatgtaaaga tagggcaaaa 660
aagctatgat ccttttgcct catctgagaa gattcagcaa caggtcagca agtaccatt 720
aatggaagag attgggttct tgggtgcttg catgagggtt tatcatgttc attccgatag 780
ctatgagaca gaaaaccagc attacggaag aagcttaaca aaagaaacta taaaggatgg 840
agtctccaga ttttttcata atgggtactg cttaagaaaa gatgctgttg ctgccagtat 900
tcagaagatt gagaaaattc tgcagtgtt tgaaaaccag aagcagctta atttttacgc 960
aagttcatta ctctttgttt atgaagggtc atctcagcca accactacaa aattgaatga 1020
cagaactttg gcagaaaagt ttttgtccaa aggacaactg tcagacacag aagtactaga 1080
gtacaataat aactttcatg tggttaagtt cacagctaat ggaaaaatag agtcttcagt 1140
gggcaaaagc ttgtccaaga tgtatgcgcg tcacaggaaa atatatacaa aaaagcatca 1200
cagtcagact tcattgaaag ttgaaaatct ggagcaagac aatgggtgga aaagcatgtc 1260
acaggaacat ttaaatggaa atgtactttc ccaactggaa aaagttttct accatcttcc 1320
cactggttgc caagagattg ctgaagtaga agtgcgaaat atagattttg ctcatgtgtt 1380
ccctagcaac acaatagatg agggatatgt ttatgggcta aagcatttaa tttctgtact 1440
tcgaagtatt ttagacaatt gaatcctctg ttgcagtctt ttttaaggggt gggccaatca 1500
taatgaagag gggcagtcac tatctgcacc tttaatgcta tgtaaaaaat ttgtattatg 1560
agtcgacatt ttatttgtct ttatactttt ggaagaatgg ttaacttttt tataatctta 1620
ctcaggaaaa ctaactattt gttcattaga aaactatgaa gaataaagaa acttaggaat 1680
gttaagcagg gaatgtggtg gtacatggct taaacatctt ttttggctca agcaaaatgc 1740
aaaccattat tcagtcatta agagttagt tagctttctg tagccaattc atgaaatctc 1800
tgtccacca gccttgacaa tgagccatat ctaaaatatt acattattag aacacctacc 1860
aaaatctcga aagcacaggt gatgtcctta gtantgctat gtatgaagtt actaaaactg 1920
gagaaaattc tacttcagaa ataagtactg tttaggtttt atattaaaag ttcagaccag 1980
catatcaaag ggtgcttctt agtgaaatga tttagaatcg tgcattccaa aagcaggttt 2040
tctctttaat ttttacatct ctctctcaaa atattatact tcatgaaaaa gacaatcgtg 2100
tggatgacaa caacaaagtc tgaaattaag ggcacactaa tgttcttacg gggtagggga 2160
gagagaattc ttttcacgca caaatatttc tttacatctt cctccggaac tg 2212

```

<210> 21
 <211> 1751
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 2378756CB1

<400> 21
 cgttcgggttc gctgtgtgtg tcgccgggtc cttgagggtc catgtgattt ttacgccagt 60
 gctgctgaac tgtgcagggt agggagctgg cacagtcga ttaattgtcc ttgggtcgag 120
 gtgtctcgtc ggaccctttg gggctcagtg gagaattaag gcagagtcac tgtaattatt 180
 tctaatacca attccaaaat agtgactctt ggacaatagt gcaattatat ggaattatgg 240
 ctgggttataa gcctgtagct attcagacat atcctatact tggtgaaaaa atcaccacaag 300
 atacactgta ctggaacaac tataagaccc ctgttcagat taaggaattt ggtgcagttt 360
 caaaagtaga cttttctcct cagcctccat ataattatgc tgtcacagct tcctcaagaa 420
 ttcacattta tggccgatac tcccaagaac ctataaaaac cttttctcga tttaaagaca 480
 cagcactactg tgctactttt cgacaagatg gtagattgct tgtggctggc agtgaagatg 540
 gtggagttca actttttgat ataagtggga gggctcccct caggcagttt gaaggccata 600
 caaaagcagt tcatacagta gattttacag ctgacaaata tcacgtgggtc tctggggctg 660
 atgattatac agttaaatga tgggatattc caaactccaa agaaattttg acatttaaag 720
 aacactctga ttatgtgagg tgtggatgtg ctagcaaac taatccggat ctctttataa 780
 caggatcata tgatcactat gtgaagatgt ttgatgcacg aacgagtgag agtggtctct 840
 ccgttgagca tgggcagcca gtggagagtg tctactttt cccctctgga ggtcttctgg 900
 tgtcagcagg aggtcgttat gttaaagtct gggacatgtt aaaaggagga caattgctag 960
 tatctttgaa aaatcatcac aaaaccgtga catgtttatg tctaagcagc tctggacagt 1020
 gggtactctc tggctcactg gataggaagg tgaaagtata cagcacaact tcctacaaag 1080
 tagtccacag ttttgattat gcagcttcaa ttttgagtct tgcccttgca catgaagatg 1140
 agacaatagt tgtaggaatg accaatggaa tactgagtgt taaacatcgg aaatctgaag 1200
 caaagaagga atcacttccc agaagaagaa ggccctgcata tcgaaccttt attaaaggaa 1260
 aaaattacat gaagcaacgg gtattttgtgc atttctcata tttgttttaa ggggtgaaatt 1320
 tgtagagta gccagcttca tccctgcaca ttactggaaa gcgtagttca catgacactt 1380
 tattttgctt ttctctacat cagaattatt tctagattgc tagattgggg acattattga 1440
 tgagtaattgg ggaacaagga ctactttatg taatactctg catcataaca taatgttata 1500
 atgctttact gcattcaaaa taatgtcttt cacatactac ctgatttgat ctaaatccca 1560
 agaagtagta gaaccaatca tgttcctgtt tggcacctga gaagactgag cgacaaataa 1620
 tttgtgattg attccggatt tcattactgg tatccacagt tgttgagact agatctcctt 1680
 atctctgaat ggcgggtggg atctccttct tctgggtccc ctgtgcgggc gcttaaaggg 1740
 aattgcgtgt g 1751

<210> 22
 <211> 4189
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 1861527CB1

<400> 22
 agcgagtccg tctgtcaggc cgcctcctct ccggccgtct gattttctac ccttcggcgc 60
 cctgctcttc ctcatgttgg catccccggc cacggagacc accgtcctca tgtcccagac 120
 tgaggccgac ctggccctgc ggccccgcgc tctcttggc accgcggggc agccccgcct 180
 cgggccccct cctcgccgag cgcgcgctt ctccgggaag gctgagcccc ggccgcgctc 240
 ttctcgtctc agccgcgcta gctcagtcga cttggggctg ctgagctctt ggtccctgcc 300
 agcctcaccg gctccggacc cccccgatcc tccggactcc gctgggtcctg gccccgcgag 360
 gagccacccg cctagctcca aagaaccccc cgagggcacg tggaccgagg gagccctgt 420
 gaaggctgcg gaagactccg cgcgtcccga gctcccgac tctgcagtgg gcccggggtc 480
 caggagccg ctaagggtcc ctgaagctgt ggccctagag cggcggcggg agcaggaaga 540
 aaaggaggac atggagaccc aggctgtggc aacgtcccc gatggccgat acctcaagtt 600
 tgacatcgag attggacgtg gctccttcaa gacggtgtat cgagggttag acaccgacac 660
 cacagtggag gtggcctggt gtgagctgca gactcggaaa ctgtctagag ctgagcggca 720

```

gcgcttctca gaggaggtgg agatgctcaa ggggctgcag caccccaaca tcgtccgctt 780
ctatgattcg tggaagtcgg tgctgagggg ccaggttttg atcgtgctgg tcaccgaact 840
catgacctcg ggcacgctca agacgtacct gaggcggttc cgggagatga agccgcgggt 900
ccttcagcgc tggagccgcc aaatcctgcg gggacttcat ttcttacct cccgggttcc 960
tcccatcctg caccgggatc tcaagtgcga caatgtcttt atcadgggac cttctggctc 1020
tgtcaaaatc ggggacctgg gcttggccac gctcaagcgc gcctcctttg ccaagagtgt 1080
catcggggacc ccggaattca tggcccccga gatgtacgag gaaaagtacg atgaggccgt 1140
ggacgtgtac gcgttcggca tgtgcatgct ggagatggcc acctctgagt acccgtactc 1200
cgagtgccag aatgccgcgc aaatctaccg caaggctact tcgggcagaa agccgaacag 1260
cttcacaaag gtgaagatac ccgaggtgaa ggagatcatt gaaggctgca tccgcacgga 1320
taagaacgag aggttcacca tccaggacct cctggcccac gccttcttcc gcgaggagcg 1380
cgggtgtgac gtggaactag cggaggagga cgacggcgag aagccggggc tcaagctctg 1440
gctgcgcagtg gaggacgcgc ggcgcggggg gcgcccacgg gacaaccagg ccatcgagtt 1500
cctgttccag ctgggcccgg acgcggccga ggaggtggca caggagatgg tggctctggg 1560
cttgggtctgt gaagccgatt accagccagt ggcccgtgca gtacgtgaac ggggttgctgc 1620
catccagcga aagcgtgaga agctgcgtaa agcaaggga ttggaggcac tcccaccaga 1680
gccaggacct ccaccagcaa ctgtgcccac ggcgcccggt ccccccagtg tcttcccccc 1740
tgagcctgag gagccagagg cagaccagca ccagcccttc cttttccgcc acgccagcta 1800
ctcatctacc acttcggatt gcgagactga tggctacctc agctcctccg gcttccctga 1860
tgctcagac cctgcccttc agccccctgg gggggtgcca tccagcctgg ctgagtcca 1920
tctctgcctg ccctcggctt ttgccctatc cattccacgt tctggccctg gaagtgactt 1980
ttcccccggg gacagctatg cctcagatgc agcttcaggc cttagcgatg tgggagaagg 2040
gatgggacaa atgaggagac cccaggggag gaatctccg cgagacccc gatcccggct 2100
gcgggtcact agtgtctcag accagaatga cagagtgggt gagtgccagc tacagacca 2160
taacagcaag atggtgacct tccgatttga tctggatggg gacagcccgg aagagattgc 2220
agctgccatg gtatataacg agttcattct gccttcggag cgagatggat ttctcagacg 2280
gattcggggag attatccagc gagtggagac cctgttgaag agagacactg gcccctatga 2340
ggctgctgaa gacaccctaa gccccagga ggagccagca ccattacctg ccctgcccgt 2400
ccccctcca gacctatcca atgaagagct ccagagcagc acctccctgg agcacaggag 2460
ctggacagcc ttctccacct cctcatcttc tcttggaaact cctttgtctc ctggaaacct 2520
atthtccccct ggaacccccca tttccccagg tcccatcttc cccatcactt ctcccccatg 2580
tcatccacagc ccctccccat tctcccccat ttcttcccag gtctcctcaa atccctctcc 2640
acaccccacc agctctccac ttccattctc ctccagcaca cccgagtthc cgggtcccact 2700
ctctcagtggt ccctggagtt ctctccccac gacttctcca cctacgttct ctcccacttg 2760
ttctcaggtc actcttagtt cccctttctt tcttccgtgc ccttccactt cttccttccc 2820
ctccaccaca gcagcccctc tcttttctct ggctagtgcc ttctcactgg ctgtgatgac 2880
tgtggcccag tccctgctgt cccctcacc tgggctcctt tcccagttct ctccagcccc 2940
tcctagtccc ctccctagcc tgcccccttc ccctcccgtt gctcctgggtg gccaggaaag 3000
cccttcaccc cacacagctg aggtggagag tgaggcctca ccacctctg ctcgccccct 3060
cccaggggaa gccaggtgg cgcccatctc tgaagaggga aagccgcagc ttgttgggcg 3120
tttccaagtg acttcatcca aggaaccggc tgagcctctt cccttgagc caacatcccc 3180
cactctctct ggttctccaa aaccttcaac ccctcagctc acttcagaga gctcagatac 3240
agaggacagt gctggaggcg ggccagagac cagggaagct ctggctgaga gcgaccgtgc 3300
agctgagggg ctgggggctg gagttgagga ggaaggagat gatgggaagg aaccccaagt 3360
tgggggcagc ccccaacccc tgagccatcc cagcccagtg tggatgaact actcctacag 3420
cagcctgtgt ttgagcagcg aggagtccga aagcagtggg gaagatgagg agttctgggc 3480
tgagctgcag agtcttcggc agaagcactt gtcagagggt gaaacactac agacactaca 3540
gaaaaaagaa attgaagatt tgtacagccg gctgggggag cagcccccac cgggtattgt 3600
ggccccagct gctatgctgt ccagccgcca gcgcgcctc tccaagggca gcttccccac 3660
ctccgcgcgc aacagcctac agcgtcttga gccccaggc cctggcatca tgcgaaggaa 3720
ctctctgagt ggcagcagca ccggctccca ggagcagcgg gcaagcaagg ggggtgacatt 3780
cgccggggat gttggcagga tgtgaattca gaacagaagc catgtatctc cccacacca 3840
gggcccacca tggagcttgt gttctcagaa tctgatgctt tctgatcaac aaaactgagc 3900
aaggaagatc ccaacactga aggggtagaa ggccaggggg gcatggagag tgcagctcca 3960
ttatagttaa gagccaaaca tatgtgaact gtttgctgtg tggaggtgtt agttctgctg 4020
cctaccatct tcatctctag cacctccctt gccagagtc aaccactaag caatcccacc 4080
caagcctgga tgcttctaga ggggcccact cccagctggg agagtgtagg ggatatgctc 4140
acaccacatt agcagcaacc aataaaaatg ctggaaacaa aaaaaaaaaa 4189

```

<210> 23

<211> 1679

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 2921356CB1

<400> 23

```
ggagcgccgg gttcgccatc cccagggcgcc ggctctgcgg ctgctgaatc ggaagccgca 60
gggaggatcc ggggaaataa agacgccgga gaatgacctc cagcgaggcc gcctgagccg 120
gggcccgcgc acagccccgc cagcccccg cagggcgac cgcagcgggc agcaggagct 180
gctcggtccc gcactctcca ggggcccccg tgggcaccag cggcgccgct gtgaacgggc 240
tgctgcacaa cggcttccat ccgcccgcag tccagccgcc gcacgtctgc agccgggggtc 300
cagtggggcg cagcgacgcg gcgccccagc gcctcccgct cctgcccggag ctgcagccgc 360
agccactgct ccctcagcat gactcccccg ccaagaaatg ccggctgcgg aggaggatgg 420
actcggggag aaagaacagg ccgccattcc catggtttgg catggacatc ggtggaacgc 480
tgggttaaatt ggtgtatttc gagccgaagg atattacagc cgaagaggag caagaggaag 540
tggagaacct gaagagcatc cgggaagtatt tgacttctaa tactgcttat gggaaaactg 600
ggatccgaga cgtccacctg gaactgaaaa acctgacct gtgtggagcg aaagggaaacc 660
tgcacttcat ccgctttccc agctgtgcta tgcacaggtt cattcagatg ggcagcgaga 720
agaacttctc tagccttcac accaccctct gtgccacagg aggcggggct ttcaaattcg 780
aagaggactt cagaatgatt gctgacctgc agctgcataa actggatgaa ctggactgtc 840
tgattcaggg cctgctttat gtcgactctg ttggcttcaa cggcaagcca gaatgttact 900
atthttgaaa tcccacaaat cctgaattgt gtcaaaaaaa gccgtactgc cttgataacc 960
cataccctat gttgctgggt aacatgggct caggtgtcag cattctagcc gtgtactcca 1020
aggacaacta taaaagagtt acagggacca gtcttgagg tggaacattc ctaggcctat 1080
gttgcttgct gactggttgt gagacctttg aagaagctct ggaaatggca gctaaaggcg 1140
acagcaccaa tgttgataaa ctggtgaagg acatttacgg aggagactat gaacgatttg 1200
gccttcaagg atctgctgta gcatcaagct ttggcaacat gatgagtaaa gaaaagcgag 1260
attccatcag caaggaagac ctgcgccggg ccacattggt caccatcacc aacaacattg 1320
gctccattgc tcggatgtgt gcgttgaatg agaacataga cagagttgtg tttgttgga 1380
atthttctcag aatcaatatg gtctccatga agctgctggc atatgccatg gatttttgg 1440
ccaaaggaca actgaaagct ctgttttttg aacatgaggg ttatttttga gccgttgggg 1500
cactgttgga actgttcaaa atgactgatg acaagtagag acgagcagtc cgaggaaaca 1560
gcctcccaa aagggacaga gaactaaaaa attgctgctg gcagaaggtg aaacgtcgct 1620
ttgggacgga agccaagcca ttatggcaga ctttaccgca cgggtgcgata tatcaatct 1679
```

<210> 24

<211> 2077

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7386170CB1

<400> 24

```
gctcccgggg cgcagcgctg acggcgggcg ggggagcgcg ccatgcccg caggaccggc 60
cccaagatgg aagggagcgg cggccgcgct cgcctcaagg cgcattacgg gggggacatc 120
ttcatcacca gcgtgggacgc cgccacgacc ttcgaggagc tctgtgagga agtgagagac 180
atgtgtcgtc tgcaccagca gcaccgctc acctcaagt ggggtggacag cgaaggtgac 240
ccttgacagg tgtcctccca gatggagctg gaagaggctt tccgcctggc ccgtcagtgc 300
agggatgaag gcctcatcat tcatgttttc ccgagcacc ctgagcagcc tggcctgcca 360
tgtccgggag aagacaaatc tatctaccgc cggggagcca gaagatggag gaagctgtac 420
cgtgccaacg gccacctctt ccaagccaag cgctttaaca gggattctgt catgccttcc 480
caagagcctc cagtagacga caagaacgag gacgccgacc ttccttccga ggagacagat 540
ggaattgctt acatttcctc atcccgaag catgacagca ttaaagacga ctcgaggac 600
cttaagccag ttatcgatgg gatggatgga atcaaatct ctcaggggct tgggctgcag 660
gactttgacc taatcagagt categggcgc gggagctacg ccaagggtct cctggtgcgg 720
ttgaagaaga atgaccaa attacgccatg aaagtgggtg agaaagagct ggtgcatgat 780
gacgaggata ttgactgggt acagacagag aagcacgtgt ttgagcaggc atccagcaac 840
cccttcttgg tcggattaca ctctgcttc cagacgacaa gtcggttgtt cctggtcatt 900
gagtacgtca acggcgggga cctgatgttc cacatgcaga ggcagaggaa gctccctgag 960
gagcacgcca ggttctacgc ggccgagatc tgcacgccc tcaacttcct gcacgagagg 1020
gggatcatct acagggacct gaagctggac aacgtcctcc tggatgcgga cgggcacatc 1080
aagctcacag actacggcat gtgcaaggaa ggccggggcc ctggtgacac aacgagcact 1140
```

```

ttctgcggaa ccccgaaatta catcgccccc gaaatcctgc ggggagagga gtacgggttc 1200
agcgtggact ggtgggcgct gggagtcctc atgttttgaga tgatggccgg gcgctccccg 1260
ttcgacatca tcaccgacaa cccggacatg aacacagagg actacctttt ccaagtgatc 1320
ctggagaagc ccatccggat ccccggttc ctgtccgtca aagcctccca tgttttaaaa 1380
ggatttttaa ataaggaccc caaagagagg ctggctgcc ggccacagac tggattttct 1440
gacatcaagt cccacgcgtt cttccgcagc atagactggg acttgctgga gaagaagcag 1500
gcgctccctc cattccagcc acagatcaca gacgactacg gtctggacaa ctttgacaca 1560
cagttcacca gcgagcccgt gcagctgacc ccagacgatg aggatgccat aaagaggatc 1620
gaccagtcag agttcgaagg ctttgagtat atcaacccat tattgctgtc caccgaggag 1680
tcggtgtgag gccgcgtgcg tctctgtcgt ggacacgcgt gattgaccct ttaactgtat 1740
ccttaaccac cgcataatgca tgccaggctg ggcacggctc cgagggcggc cagggacaga 1800
cgcttgccgc gagaccgcag agggaaagcgt cagcgggcgc tgctgggagc agaacagtcc 1860
ctcacacctg ggcccgggca ggccagcttc gtgctggagg aacttgctgc tgtgcctgcg 1920
tcgcgggcga tccgcgggga ccctgccgag ggggctgtca tgcggtttcc aaggtgcaca 1980
ttttccacgg aaacagaact cgatgcactg acctgctccg ccaggaaagt gagcgtgtag 2040
cgtcctgagg aataaaatgt tccgatgatg tgaaaaa 2077

```

<210> 25

<211> 3498

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7481206CB1

<400> 25

```

tagggtgaaa gccgccctca cgctcacttc cggcaggcgc gcctcccgac agtgtcgtgc 60
ggggggcgctg gcccggcatt gctgacaggc ggccccgggg gcggtggcca aggcggcgac 120
cggagcgcga tggcgggggc ggcgggactc acggcagaag tgagctggaa ggtcttgag 180
cgaagagctc ggaccaagcg ctcaggctca gtttatgaac ctcttaaaag cattaatctt 240
ccaagacctg ataatgaaac tctctgggat aagttggacc attattacag aattgtcaag 300
tcaacattgc tgctgtatca aagtccaact accggtctct tcccactaa aacatgcggt 360
ggtgaccaga aggccaagat ccaggacagc ctatactgcg ctgctggggc ctgggctttg 420
gctcttgcat acaggcgaat tgatgatgac aagggaagga cccatgagct ggagcactca 480
gctataaaat gcatgagagg aattctctac tgctatatgc gtcaggccga taaggctccag 540
cagtttaagc aggatccacg cccaacaaca tgtcttcaact ctgttttcaa tgtgcataca 600
ggagatgagt tgctttccta tgaggaatat ggtcatcttc agataaatgc agtgtcactt 660
tatctccttt accttggtga aatgatttcc tcaggactcc agattatcta caacactgat 720
gaggtctctt ttattcaaaa ccttgatatt tgtgtggaaa gagtttaccg tgtgcctgac 780
tttggtgtct gggaaagagg aagcaaatat aataatggca gcacagagct acattcgagc 840
tcggttggtt tagcaaaagc agctctagaa gcaattaatg gattcaacct ttttggaac 900
cagggctgtt cgtggctcagt tataattgtg gatctcgatg ctcaaatcg caacaggcaa 960
actttgtgct cgctgttacc cagagaatca agatcacata atacagatgc tgccctgctc 1020
ccctgcatca gttatcctgc atttgccctg gatgatgaag ttcttttttag ccagacactt 1080
gataaagtgg ttagaaaatt aaaaggaaaa tatggattta aacgtttctt gagagatggg 1140
tatagaacat cattggaaga tcccaacaga tgctactaca agccagctga aattaagcta 1200
tttgatggca ttgaatgtga atttcccata tttttccttt atatgatgat tgatggagtt 1260
tttagaggca atcctaagca agtacaggaa tatcaggatc ttttgactcc agtacttcat 1320
cataccacag aaggatatcc tgttgtacca aagtactatt atgtgccagc tgactttgta 1380
gaatatgaaa aaaataaccc tggtagtcaa aaacgatttc ctagcaactg tggccgtgat 1440
ggaaaactgt ttctttgggg acaagcactt tatatcatcg caaaactcct ggctgatgaa 1500
cttattagtc ctaaagacat tgatcctgtc cagcgctatg tcccactaaa ggatcaacgt 1560
aacgtgagca tgagggttttc caatcagggc ccactggaaa atgacttggg agttcatgtg 1620
gcacttatag cagaaagcca aagacttcaa gtttttctga acacatatgg tattcaaact 1680
caaactcctc aacaagtaga acccattcag atatggcctc agcaggagct tgtgaaagct 1740
tatttgacgc tgggtatcaa tgaaaagtta ggactctctg gaaggccaga caggccatt 1800
ggctgcctcg ggacatcaaa gatttatcgc attctaggaa agactgtggg ttgttaccgg 1860
attattttcg acctaaagtga tttctacatg tctcaggatg ttttcctgct gatagatgac 1920
ataaagaatg cgctgcagtt cattaaacaa tattggaaaa tgcattggacg tccacttttc 1980
cttggtctca tccgggaaga caatataaga ggtagccggg tcaaccccat attagatatg 2040
ctggcagccc ttaaaaaagg aataattgga ggagtcaaag ttcatgtgga tcgtctacag 2100
acactaatat ctggagctgt ggtagaacia cttgatttcc tacgaatcag tgacacagaa 2160

```



```

gagcttccag aatttaagag ttttgaggaa ctagaacctc ccaaacattc aaaagtcaaa 2220
cggcaaagca gcacccctag tgctcctgaa ctgggacagc agccggatgt caacattagt 2280
gaatggaagg acaaaccac ccacgaaatt cttcaaaaac tgaatgattg cagttgtctg 2340
gctagccaag ccattcctgct ggggtatactg ctcaaaagag aaggccccaa cttcatcaca 2400
aaggaaggta ccgtttctga tcacattgag agagtctata gaagagctgg cagccaaaaa 2460
ctttggtcgg ttgtacgccg tgcagcaagt cttttaagta aagtagtgga cagcctggcc 2520
ccatccatta ctaatgtttt agtgcagggc aaacaggtaa ctctgggtgc ctttgggcat 2580
gaagaagaag ttatctctaa tcctttgtct ccaagagtga ttcaaaacat catctattat 2640
aagtgtaca cccatgatga gagggaagcg gtcattcagc aagaactggc catccatatt 2700
ggctggatca tctccaataa ccctgagtta ttcagtggca tgctgaaaat acgaatcggg 2760
tggatcatcc atgccatgga gtatgaactt cagatccgtg gcggagacaa gccagccttg 2820
gacttgatc agctgtcacc tagtgaagtt aaacagcttc tgctggatat tctgcagcct 2880
caacagaatg gaagatgttg gctgaacagg cgtcagatcg atgggtcttt gaatagaact 2940
cccaccgggt tctatgaccg agtgtggcag attctggagc gcacgcccac tgggatcatt 3000
gttgctggga agcatttgcc tcagcaacca accctgtcag atatgaccat gtatgagatg 3060
aatttctctc tccttggttg agacacgttg ggaaatattg accagccaca gtacagacag 3120
atcgtttag agttacttat ggttgatcc attgtactgg aaagaaacc cgagctagaa 3180
tttcaagaca aagtagatct agacagactg gtcaaagaag catttaatga atttcaaaaa 3240
gatcagagtc ggctaaagga aattgaaaaa caagatgaca tgacttcctt ttacaacact 3300
cctcccctgg gaaaaagagg aacatgcagc tatttgacaa aggcggtgat gaatctgctg 3360
ctggaaggag aagtcaagcc aaacaatgat gaccctgtc tgattagcta gtggggaagg 3420
tgtaggaagc tctgttgaga cacatgttct gaagtgtgtt gtgtttcatg ttcagcttat 3480
caaggcagcc attaatat 3498

```

<210> 26

<211> 5540

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7503117CB1

<400> 26

```

ccgcttcgcc agacagccag cggccggcgg caggccgggc catgagcggc aggggcccggg 60
ccgggcctcg ctgaccctgg ctccgcgcgg cagcttcccc agtttccgct ccggtctctc 120
ggcatgagag tccgcccggg cccggggctg cggctgcccc agaccgcgcg cacgctggcg 180
cgctccgggc ccgcgagacc gcgggtgctga tacctgcgcc gcactgcgcc gccgcgccgt 240
ccgctgtgtg ccccgggggc gcggccatgg aggtggtggg tgacttcgag tacagcaaga 300
gggatctcgt gggacacggg gccttcgccg tggctctccg ggggcggcac cgccagaaaa 360
ctgattggga ggtagctatt aaaagtatta ataaaaagaa cttgtcaaaa tcacaaatac 420
tgcttggaag ggaaattaaa atcttaaaag aacttcagca tgaaaatatt gtagcactct 480
atgatgttca ggaattaccc aactctgtct ttttgggtgat ggagtattgc aatggtggag 540
acctcgcaga ttatttgcaa gcgaaagggg ctctcagtga agacacgate agagtgtttc 600
tgcatcagat tgctgtctgc atgcgaatcc tgcacagcaa aggaatcatc cacagagatc 660
tcaaaccaca gaacatcttg ctgtcctatg ccaatcgcag aaaatcaagt gtcagtggta 720
ttcgcaccaa aatagcggat tttgggtttt ctcgttacct acatagtaac atgatggctg 780
caacactgtg tggatccccg atgtacatgg ctctgaggt tattatgtct caacattatg 840
atgctaaggc tgacttgtgg agcataggaa cagtgatata ccaatgccta gttggaaaac 900
caccttttca ggccaatagt cctcaagact taaggatgtt ttatgaaaaa aacaggagct 960
taatgcctag tattcccaga gaaacatcac cttatttggc taatctcctt ttgggttttg 1020
ttcagagaaa ccaaaaagat agaattggact ttgaagcgtt ttttagccat ccttttcttg 1080
agcaaggctc agtaaaaaaa tcttgcccag ttccagtgcc catgtattct ggttctgtct 1140
ctggaagctc ctgtggcagc tctccatctt gtcgttttgc ttctccacca tcccttccag 1200
atatgcagca tattcaggaa gaaaacttat cttccccacc attgggtcct cccaactatc 1260
tacaagtttc caaagattct gccagtacta gtagcaagaa ctcttcttgt gacacggatg 1320
actttgtttt ggtgccacac aacatctcgt cagaccactc atgtgatatg ccaatgggga 1380
ctgctggcag acgtgcttca aatgaattct tgggtgtgtg agggcagtg cagcctactg 1440
tgtcacctca cagcgaaaca gcaccaattc cagttcctac tcaaataagg aattatcagc 1500
gcatagagca gaatcttaca tctactgcca gctcaggcac aaatgtacat ggttctccaa 1560
gatctgcagt ggtacgaagg tccaacacca gcccctggg ctctctccgg ccgggatcat 1620
gctcccagc accagcagac acagcacaga cagttggacg aaggctctcc actgggtctt 1680
ctaggcctta ctcaccttcc cctttggttg gtaccattcc tgagcaattc agtcagtgtc 1740

```

gctgtgggca	tcctcagggc	catgactcca	ggagtagaaa	ctcctcaggt	tctccagtgc	1800
cacaagctca	gtccccacag	tctctcttat	cgggtgctag	actgcagagc	gccccacccc	1860
tcactgacat	ctatcagaac	aagcagaagc	tcagaaaaca	gcactctgac	cccgtgtgcc	1920
catcccatac	tggggctggg	tacagctact	cgcctcagcc	cagtcggcct	ggcagccttg	1980
gaactttctc	caccaagcac	ttggggtcct	ctccacggag	ttctgactgg	ttcttttaaaa	2040
ctccttttgc	aacaatcatt	ggctctccta	ctaagaccac	agctcctttc	aaaatcccta	2100
aaactcaagc	atcttccaac	ctgttagcct	tggttactcg	tcattgggct	gctgaagaac	2160
agtcgaaaga	tgggaatgag	ccacgggaat	gtgcccattg	cctcttagtg	caaggaagtg	2220
agaggcagcg	ggccgagcag	cagagcaagg	cagtgttttg	cagatctgtc	agtaccggga	2280
agttatcaga	tcaacaagga	aagactccta	tatgtcgaca	tcagggcagc	acagacagtt	2340
taaatacaga	acgaccaatg	gatatagggt	ctcctccaca	cagtgcggca	gccccactt	2400
gtacccacat	gttccttcga	acaagaacaa	cctcagtggg	gcccagcaac	tccgggggct	2460
ctcttttgtc	catgagtggc	cgcgtgtgcg	tgggggtccc	gcctggccca	ggcttcggct	2520
cttccccctc	aggagcagag	gcagctccca	gcctgagata	cgtgccttac	gggtgcttcac	2580
cccccagcct	agaggggctc	atcacctttg	aagccccctg	actgccggag	gagacgctga	2640
tggagcggga	acacacagac	accttacgcc	atctgaatgt	gatgctgatg	ttcactgagt	2700
gtgtgctgga	cctgacagcc	atgaggggag	gaaaccctga	gctgtgcaca	tctgctgtgt	2760
ccttgtacca	gatccaggag	agtgtggtgg	tggaccagat	cagtcagctg	agcaaagact	2820
ggggggcggg	ggagcagctg	gtgtttgtaca	tgaagcagc	acagctgctt	gcggcttctc	2880
tgcactcttg	caaagcccag	atcaagtcct	ggaaactgag	cccatccaca	gctgtgaaac	2940
aagtgtgcaa	gaatctgaac	gaacgatata	aattctgcat	caccatgtgc	aagaaactta	3000
cagaaaagct	gaatcgattc	ttctctgaca	aacagagggt	tattgatgaa	atcaacagtg	3060
tgactgcaga	gaaactcatc	tataattgtg	ctgtagaaat	ggttcagctc	gcagccctgg	3120
atgagatggt	tcagcagacc	gaagatattg	tttatcgcta	tcataaggca	gcccttcttt	3180
tggaaggcct	aagtaggatt	ctacaggacc	ctgcagatat	tgaaaatgtg	cataaatata	3240
aatgtagtat	tgagagaaga	ctgtcggcgc	tctgccatag	caccgcaacc	gtgtgagcag	3300
caggctcatc	ccgtggaccg	gtggtgggaa	cgtgaggtga	tgcctttggg	attacagctt	3360
gagttctgtc	accccatccc	caggaaactg	tagcttctta	actggtgact	accaaagaac	3420
aagcagtgat	ttgaaaaagg	aaaaacaatc	caaaaactac	atattttgtg	gaaatctgcc	3480
ttattggaga	aaatcacccct	ttcccttttt	ctttgtagaa	gcaggagcaa	gagtgttttg	3540
ctcccagttt	ggacttggtg	aataaatgta	ccttagaact	aggataatcg	gtacagttat	3600
tcttaaagat	aattaaaaat	gaaacaaagt	gagtgtctgt	cactgggttc	atcagagcag	3660
tgtgtgaaat	tccatgtgtt	tgctgaggtg	taaaggtaaa	tgtattcacc	cctcatccag	3720
gcagtttgat	atttggagta	agttttgttta	aatctgagca	tgcactctta	aacagctcag	3780
gaagaaatag	cttaagaaga	agtgaacat	ggatcttgga	agaaattttg	aaatcttcaa	3840
tttgatccta	atatggatac	atgttaatct	tccaaaatct	ttcatattgc	actaatttat	3900
taaaacaact	gtgtattgga	ttttgtaatt	taactaaggc	acaatggact	tgttttaaat	3960
attttacttg	attgtatata	tagacccttt	ccagaattca	catgtaatct	ccagtgaact	4020
tttaagtgg	taaaacttgt	attcatgtga	acctttgcac	attttttttt	ttttacttct	4080
ttatctacac	ctacagattt	tctcagtaat	gtttttgtta	gcttttggtt	ccatttttta	4140
ttgtgcatgc	agaatgtaca	ttgatgcctg	tgaccttagg	tttattaaag	gctaggttta	4200
tttgggcagt	attagaacaa	aaatcatgga	tcaagagata	ctcttgataa	tttgaatagg	4260
gccccaaaca	agttggtgac	ctaaaggcct	gttagtgatg	tggagtccct	acatgcagtg	4320
agtggaataa	gaagtctgtt	ttctcttagg	aaaatgggca	gctgtcttct	gcctaattgtg	4380
tatttttcat	gttaattctg	acagttcacc	aaatagctag	tcattggagaa	tgcaggcagt	4440
taacttaata	tccctccagg	aatgggttct	acgttgtgta	ttatttggtt	tcttttactt	4500
acctgcttga	atacttgaat	aaaccattca	ccaattttta	tccttttatt	ttaatccttt	4560
tacataaaat	aatctgaact	ctttgacaaa	ttgcacagag	ctcttttgga	ttaatctaat	4620
tttaatgtac	tgataaaaaa	aaacatgggt	gtcctttact	ttgacaaagt	aatgtaattt	4680
ttaccttatt	tatctgtatg	aaattccagt	agtttaattt	aacattttatt	tatatgacgt	4740
ttgtattttt	aggtctttta	tacagtgttt	ctacctctca	tttgtaactg	catgcattat	4800
tcttgaaact	aggtaaaact	cactgaattg	ttgtgtaata	gcctttttat	tattgcctgt	4860
acaaatgtat	attaaggtaa	aataaaaact	acaaagtgtt	tctagggtgt	agctgggtac	4920
atattaagtg	gcttggttgag	ccaggtactt	ccttagtgag	tttagagact	tggccatgaa	4980
tatcctttgt	cctgccccag	gatttagatc	ttggctactg	tcattgcaggc	ttccagggaac	5040
atagactggt	ttacctccac	aacctatttt	gttattagtg	atactttatt	ttatataata	5100
tttttttatc	acagtgaat	ttcattcatg	ttctttcagt	tatcacctgt	gttatctcag	5160
ttgtagggtt	attctatcct	ctcctcttcc	tctccatttt	cttttttaac	acaggatgaa	5220
acaggttcag	agaggggaag	tgattggcct	aaagtcagga	actaggcaag	tgggtcaagcc	5280
atgcttttgt	actttcaagt	taattcttct	tgttcttgta	tattaaagggt	cttggggtag	5340
atgggtgtgt	tgaaacagtg	aagtcctcaac	agcagaaaag	aacaaaatgt	aaattcatga	5400
ataatgggtc	tggttatact	tccattatca	aggctaatta	agagattttg	ccttgagtat	5460
agcaataata	aacaaatgct	ttatgttttc	ctgaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	5520

aaaaaaaaaa aaaaaaaaaa

5540

<210> 27

<211> 1785

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7506911CB1

<400> 27

```

atgaagctta taaatggcaa aaagcaaagg actacttctg atgaagaaag tggcataatt 60
aaagctgagg acctcacgcc tgtaatcccc ccactttggg aggctgaggc ggggtggatca 120
cgagcattcc catggttttg catggacatc ggtggaacgc tggttaaatt ggtgtatttc 180
gagccgaagg atattacagc cgaagaggag caagaggaaag tggagaacct gaagagcatc 240
cggaagtatt tgacttctaa tactgcttat gggaaaactg ggatccgaga cgtccacctg 300
gaactgaaaa acctgaccat gtgtggacgc aaagggaacc tgcacttcat ccgctttccc 360
agctgtgcta tgcacagggt cattcagatg ggcagcgaga agaacttctc tagccttcac 420
accaccctct gtgccacagg aggcggggct ttcaaattcg aagaggactt cagaatgatt 480
gctgacctgc agctgcataa actggatgaa ctggactgtc tgattcaggg cctgctttat 540
gtcgactctg ttggcttcaa cggcaagcca gaatgttact attttgaaaa tcccacaaat 600
cctgaattgt gtcaaaaaaa gccgtactgc cttgataacc cataccctat gttgctgggt 660
aacatgggct caggtgtcag cattctagcc gtgtactcca aggacaacta taaaagagtt 720
acagggacca gtcttgagg tggaacattc ctaggcctat gttgcttgct gactggttgt 780
gagacctttg aagaagctct ggaaatggca gctaaaggcg acagcaccaa tgttgataaa 840
ctgggtgaagg acatttacgg aggagactat gaacgatttg gccttcaagg atctgctgta 900
gcatcaagct ttggcaacat gatgagtaaa gaaaagcgag attccatcag caaggaagac 960
ctcgcgccgg ccacattggg caccatcacc aacaacattg gctccattgc tcggatgtgc 1020
gcgttgaatg agaacataga cagagttgtg ttgttgga attttctcag aatcaatatg 1080
gtctccatga agctgctggc atatgccatg gatttttggg ccaaaggaca actgaaagct 1140
ctgttttttg aacatgaggg ttatttttgga gccgttgggg cactgttggg actgttcaaa 1200
atgactgatg acaagtagag acgagcagtg gaggaaacag cctcccaaaa ggacagagaa 1260
ctaaaaaatt gctgctggag aaggtgaaag tcgctttggg acggaagcca agccattatg 1320
gcagatgaac ctgctggatt tgtaaataat ttaaaatcct tcagatgat cttttactct 1380
taggttttga gctaatgatt caaaacgggg gaataataaa ggtttttttt ctgtatactg 1440
tattttttta aaaaaatggg gcagcgtggc caaacctacc aattgtatgc attaactttg 1500
aaaagttgtt tgatgtttta gaaggacctg atatgtaagc gctgggtcatt tttcttctgg 1560
ggtttactga tcagtgtggg gatttttaact tcatttagta attactctag gagattttac 1620
cttgacttat atttttcatg acgtttcatg atttgctgtt ggtttcaaat gaaactacaa 1680
atctggcatg ttttactgtg aacacttttg ttatttgttt tgtacccttt tttgtcttgt 1740
ttttctgttt tagttgtctt ctgaaaaaag agttgttccc tctgt 1785

```

<210> 28

<211> 1615

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7510809CB1

<400> 28

```

cgttcgggtc gctgtgtgtg tcgccggctc cttgaggggc catgtgattt ttacgccagt 60
gctgctgaac tgtgcagggt agggagctgg cacagtcgga ttaattgtcc ttgggtcgag 120
gtgtctcgtc ggaccttttg gggctcagtg gagaattaag gcagagtcac tgtaattatt 180
tctaatacca attccaaaat agtgactctt ggacaatagt gcaattatat ggaattatgg 240
ctggttataa gcctgtagct attcagacat atcctatact tggtgaaaaa atcacccaag 300
atacactgta ctggaacaac tataagaccc ctgttcagat taaggaattt ggtgcagttt 360
caaaagtaga cttttctcct cagcctccat ataattatgc tgtcacagct tcctcaagaa 420
ttcacattta tggccgatac tccaagaac ctataaaaac cttttctcga tttaaagaca 480
cagcatactg tgctactttt cgacaagatg gtagattgct tgtggctggc agtgaagatg 540
gtggagttca actttttgat ataagtggga gggctcccct caggcagttt gaaggccata 600

```

caaaagcagt	tcatacagta	gatttttacag	ctgacaaata	tcacgtgggc	tctggggctg	660
atgattatac	agttaaatta	tgggatattc	caaactccaa	agaaattttg	acatttaaag	720
aacactctga	ttatgtgagg	tgtggatgtg	ctagcaaact	taatccggat	ctctttataa	780
caggatcata	tgatcatact	gtgaagatgt	ttgatgcacg	aacgagtgag	agtgttctct	840
ccgttgagca	tgggcagcca	gtggagagtg	tcctactttt	cccctctgga	ggtcttctgg	900
tgtcagcagg	aaggtgaaag	tatacagcac	aacttcctac	aaagtagtcc	acagttttga	960
ttatgcagct	tcaattttga	gtcttgccct	tgcacatgaa	gatgagacaa	tagttgtagg	1020
aatgaccaat	ggaatactga	gtgttaaaca	tcggaaatct	gaagcaaaga	aggaatcact	1080
tcccagaaga	agaaggcctg	catatcgaac	ctttattaaa	ggaaaaaatt	acatgaagca	1140
acgggtatth	gtgcatttct	catatttggt	taaaggggtg	aatttggttag	agtagccagc	1200
ttcatccctg	cacattactg	gaaagcgtag	ttcacatgac	actttatttt	gcttttctct	1260
acatcagaat	tatttctaga	ttgctagatt	ggggacatta	ttgatgagta	atggggaaca	1320
aggactactt	tatgtaatac	tctgcatcat	aacataatgt	tataatgctt	tactgcattc	1380
aaaataatgt	ctttcacata	ctacctgatt	tgatctaaat	cccaagaagt	agtagaacca	1440
atcatgttcc	tgtttggcac	ctgagaagac	tgagcgacaa	ataatttggtg	attgattccg	1500
gatttcatta	ctggtatcca	cagttggtga	gactagatct	ccttatctct	gaatggcggg	1560
tgggatctcc	ttcttctggt	ccccctgtgc	gggcgcttaa	aggggaattgc	gtgtg	1615